Perception And Representation of Faces

Abhandlung
zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der Universität Zürich

Vorgelegt von
Adrian Schwaninger
von Zürich

Angenommen auf Antrag von
Prof. Dr. W. Marx

Zürich 2003

Zentralstelle der Studentenschaft

Contents

1 Summary

Face recognition is one of the most important abilities of visual cognition. Whereas different object classes can be distinguished by pronounced differences in colour, texture, shape and part structure, each face contains the same parts in a very similar position. In order to recognize an individual, subtle differences between features and spatial relations need to be detected. Features (or component information) refer to the parts of the face such as the eyes, nose, mouth, etc. Configural information refers to the spatial interrelationship of the components, such as the inter-eye distance or the eye-mouth distance. In Experiments 1-4 it was shown that detecting component changes is relatively independent on orientation whereas the detection of configural information is strongly impaired when faces are rotated away from the normal upright position. The results are discussed in relation to current hypotheses for the face inversion effect. In Experiments 5 and 6 the perception of configural information was investigated as opposed to recognition based on configural processing. A new and large perceptual illusion was found. Configural information such as the inter-eye distance or the eye-mouth distance is overestimated by 15-40%. These effects remain stable across changes of orientation, which shows that perception and recognition of faces based on configural information are qualitatively different processes. In Experiments 7 and 8 a psychophysical method was developed to investigate whether truly separate systems exist for representing component and configural information. Previously learnt faces could be recognized when they were presented scrambled or low pass filtered. These manipulations selectively eliminate either configural information (by scrambling the components) or component information (by low pass filtering and gray-scaling faces). It was also shown that when faces are familiar, they can be recognized better, but there is no qualitative processing change, i.e. both processing modes become more effective to a similar extent. In Experiment 9 a repetition priming study was conducted that allowed investigating whether the outputs of component and configural processing are integrated into a holistic face identification unit. This seemed to be the case because similar effects of priming were found from scrambled to blurred recognition and vice versa. In Experiment 10, face recognition was investigated with regard to changes in viewpoint. Systematic effects were found that are consistent with view interpolation models. The results could be modelled using a recognition algorithm based on key-frames, which stores faces as a collection of temporally associated views (Experiment 11). In chapter seven the computer model was extended to explain component and configural processing by humans. The computer simulations showed a very high similarity to human data, indicating that the concept of two separate routes that are integrated for recognition is not only psychophysically but also computationally plausible. In the last chapter a summary of all experiments is presented and discussed in relation to other studies.

2 Component and configural processing of rotated faces

2.1 Abstract

The effect of orientation upon face recognition was explored by selectively altering facial components (eyes and mouth) or by changing configural information (distances between components). Regardless of type of change, a linear increase in reaction time of same-different judgements was revealed when the faces were rotated away from upright. The analyses of error scores indicated that the detection of altered components was only little affected by orientation, while orientation had a detrimental effect upon the detection of configural changes. The results of four experiments support the view that 1) the face inversion effect is due to capacity limitations of an orientation normalization process, and 2) that face recognition relies on separate representations for configural and component information.

2.2 Introduction

It was well known already by painters and Gestalt psychologists that face processing is highly dependent on orientation (e.g., Köhler, 1940). Using a forced choice recognition paradigm, Yin (1969) revealed that face recognition is disproportionately affected by inversion when compared to the recognition of other mono-oriented objects such as airplanes, houses, and stick figures of men in motion. This finding has been referred to as the face inversion effect. Subsequently, several studies have provided further evidence for the existence and robustness of this phenomenon (Ellis, 1975; Goldstein & Chance, 1981; Scapinello & Yarmey, 1970; Yarmey, 1971). In the last thirty years, at least five different hypotheses have been proposed to account for the fact that the processing of faces is so orientation sensitive.

First, the component-configural hypothesis is based on a qualitative distinction between component and configural information. The term component (or componential, piecemal, featural) information has been referred to facial elements, which are perceived as distinct parts of the whole such as the eyes, mouth, nose or ears. Researchers have used at least two different ways of manipulating component information. Rhodes, Brake, and Atkinson (1993), Sergent (1984), Tanaka and Farah (1993), and Tanaka and Sengco (1997) replaced the individual components whereas Searcy and Bartlett (1996), Leder and Bruce (1998), and Murray, Yong, & Rhodes (2000) altered color and shape attributes (e.g., darkening and whitening of pupils, elongating or blackening of teeth). The term configural information has been referred to as the ”spatial interrelationship of facial features” (Bruce, 1988, p.38). Similar meanings convey the terms configurational, spatial-relational, and second order relational information. In practice, quite different manipulations have been used to change configural information. One widely accepted method consists in altering the distance between components (Leder & Bruce, 1998; Murray et al., 2000; Searcy & Bartlett, 1996; Sergent, 1984; Tanaka & Sengco, 1997). The Thatcher illusion as illustrated by Thompson (1980), i.e. inverting the eyes and mouth within the otherwise upright facial context, has also been conceived as a configural manipulation (Bartlett &

Searcy, 1993; Diamond & Carey, 1986; Murray, et al., 2000; Stevenage, 1995). Young, Hellawell, & Hay (1987) claimed to have changed the configural information by misaligning the top and bottom halves of face composites. These three different forms of configural manipulations have in common that their detection is much more impaired by inversion than the detection of component alterations. Such a differential effect of inversion is predicted by the component-configural hypothesis. According to this view, component and configural information in faces are processed by two separate mechanisms. The processing of configural information is strongly affected by orientation while the processing of component information should be affected much less - if at all - when faces are disoriented. Face recognition is so orientation sensitive because it relies much more on the detection of subtle (mainly configural) differences than basic level object recognition. Therefore, inversion impairs the recognition of faces much more than the recognition of other objects (e.g., Bartlett & Searcy, 1993; Carey & Diamond, 1977; Diamond & Carey, 1986; Searcy & Bartlett, 1996; Sergent, 1984).

A second hypothesis is based on the assumption of "holistic" processing. According to Farah, Drain, & Tanaka (1995) "face perception is holistic and the perception of holistically represented complex patterns is orientation sensitive" (p. 633). Hence, faces are stored holistically, i.e. as unparsed perceptual wholes, in which individual parts or components are not explicitly represented. Such holistic processing is impaired when faces are substantially rotated away from their upright orientation, which results in the face inversion effect (Farah et al., 1995; Farah, Wilson, Tanaka, & Drain, 1998; Tanaka & Farah, 1991; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). A purely holistic view of face processing implies that faces are encoded and represented as undecomposed wholes without any explicit part representations. Therefore, a distinction between components and configurations becomes superfluous thus contrasting the predictions of the component-configural hypothesis.

A third hypothesis for the effect of orientation upon face processing derives from the *multidimensional space framework* (MDF). According to Valentine (1991), face recognition can be envisaged as a process of encoding a face in an n-dimensional face space. Its dimensions, which are acquired via perceptual learning, represent important physiognomic features. In this view, perceptual errors are produced by impaired encoding conditions (e.g., viewing time, stimulus quality etc.) and by the proximity of neighboring faces, which compete for matching the input face. The MDF provides a useful heuristic for a variety of effects such as the own-race bias, effects of distinctiveness, and the caricature advantage (Valentine, 1991, 1995). According to Valentine (1991) "presenting a face upside-down is an example of one experimental manipulation that would make the encoding conditions difficult, leading to a large error associated with the location in the multidimensional space derived from a stimulus face" (p. 172). However, the dimensions of the face space are not specified and this hypothesis does therefore not yield decisive conclusions regarding a selective effect of orientation upon certain types of facial information.

Goldstein and Chance (1980) have suggested a fourth hypothesis for the face inversion effect. According to their view, the fact that inverted faces are so difficult to recognize is attributable to the development of a "face schema". The ability to process faces is assumed to improve depending on the amount of exposure. They suggest that this improvement is attained at the expense of flexibility. Faces are usually seen upright and the performance to recognize upright faces improves with age. However, as a consequence of increased experience, the ability to deal with unusual orientations would actually decline through development. This has been supported by developmental studies of face recognition (for a discussion see Carey, 1992; Ellis, 1992; Johnston & Ellis, 1995). Diamond and Carey (1986) revealed that participants were affected by inversion when tested with human faces but not when dog profiles had to be recognized. In contrast, dog experts such as experienced dog show judges and breeders showed an effect of inversion on their recognition of dog profiles. Thus, based on the schema hypothesis, one would assume that a large amount of exposure has resulted in a "dog schema", which is orientation sensitive. However, Goldstein and Chance (1980) do not elaborate the actual mechanism by which a face schema is used. Therefore, it is not possible to delineate further differences with regard to the encoding and representation of faces except for the relation between experience and orientation sensitivity.

The fifth hypothesis is based on the assumption that disoriented faces have to be mentally rotated in order to recognize them (Rock, 1973, 1974, 1988). Because faces are such complex stimuli, it is not possible to rotate all the features simultaneously, which makes it difficult to detect configural changes. In contrast, mentally rotating features is sufficient to recognize part changes. As pointed out by Valentine & Bruce (1988) the mental rotation hypothesis makes similar predictions as the component configural hypothesis. Both accounts predict a small effect of orientation upon error scores for detecting component changes but high error scores for detecting configural changes. However, the component-configural hypothesis and the mental rotation hypothesis differ in their predictions on response times (RTs). The component-configural hypothesis assumes two separate mechanisms for processing component and configural information. Only the configural processing system is meant to be strongly impaired when faces are rotated. Therefore, a strong orientation effect is predicted for the detection of configural changes which applies to error scores and RTs. Because the system for processing component information is much less orientation sensitive, the detection of component alterations should not be remarkably affected by changes of orientation. In contrast, the mental rotation hypothesis predicts an increase of RT with increasing orientation for the detection of component changes as well as for processing configural alterations because in both cases the system is limited to mentally rotate facial features. Moreover, a differential effect for same vs. different trials is predicted. In a different trial the detection process can be stopped as soon as one component change has been detected. In a same trial several components would have to be mentally rotated to test whether they are the unchanged. Thus RTs should be higher and steeper for same trials than for different trials and this should be more evident for detecting component changes.

The holistic hypothesis claims that processing facial information as a whole is impaired when faces are substantially rotated away from the vertical. Also the mental rotation hypothesis also assumes that processing rotated faces as a whole is not possible because faces are too complex stimuli and thus overtax a mental rotation mechanism. Despite of this accordance, the mental rotation hypothesis and the holistic hypothesis differ with regard to their predictions on the processing of component and configural information in rotated faces. As pointed out by Searcy and Bartlett (1996) the holistic hypothesis assumes that rotating a face disrupts the processing of what is nominally

described as configural *and* component information. In contrast, the mental rotation hypothesis assumes that rotated faces are processed by mentally rotating facial features. Consequently, relatively small orientation effects are predicted for error scores when component changes have to be detected. However, strong orientation effects upon error scores for detecting configural changes are predicted because rotating facial features one after the other makes it difficult to cope with the spatial relationship of the parts (for a similar view see Valentine & Bruce, 1988).
Finally, the MDF hypothesis claims that inversion leads to encoding errors that make face recognition less accurate and tentatively also slower. However, the dimensions of the face space are not specified and therefore the MDF hypothesis does not predict any differential effect in detecting configural or component changes. The same limitation applies to the schema hypothesis as well.
Table 1 summarizes the different predictions emerging from the five hypotheses mentioned above.

	Predictions	
Hypothesis	*Error scores*	*Reaction times*
Component-configural	Components → Configuration ↗	Components → Configuration ↗
Mental rotation	Components → Configuration ↗	Components ↗ Configuration ↗
Holistic	Components ↗ Configuration ↗	Components ↗ Configuration ↗
MDF	↗ (No separate predictions)	↗ (No separate predictions)
Schema	↗ (No separate predictions)	↗ (No separate predictions)

Table 1 Predicted effects of orientation on error scores and reaction times for the detection of component and configural alterations. *Note.* Predicted increases with increasing orientation are indicated by ↗. The symbol → means no or small effect of orientation is predicted.

The aim of the present study is to contribute to a resolution of the ongoing debate between different theoretical accounts for the face inversion effect. To this end, several aspects should be taken into account. Previous studies on the processing of component and configural information examined the processing of upright and upside down faces and were concerned mainly with error rates (e.g., Bartlett & Searcy, 1993; Leder & Bruce, 1998; Rhodes et al., 1993; Searcy & Bartlett, 1996; Sergent, 1984). The present study includes several angles of rotation between upright and inverted, which allows further investigating the underlying mechanisms of face recognition. If, as suggested by the component-configural hypothesis, upright and upside down faces were processed differently, an incremental variation of orientation might reveal a discontinuity in the processing strategy. In fact, Cochran, Pick, and Pick (1983) revealed a significant nonlinear component in one particular task, although it must be noted that highly schematic profile faces were used. In a recent study by Murray et al. (2000) a discontinuity in the function relating bizarreness and orientation was revealed between 90° and 120°, which was found for thatcherized faces and faces in which configural changes were induced by changing the relative position of the eyes and mouth. Interestingly, the bizarreness ratings of unaltered or component-distortion faces (teeth blackened and eyes whitened) showed only a linear trend. Note that investigations on matching and recognizing faces revealed only linear effects of orientation (Bruyer, Galvez, & Prairial, 1993; Rock, 1973, 1974; Sergent & Corballis, 1989; Valentine & Bruce, 1988). Even though a non-linear relationship has not been supported empirically in these latter studies, it may not be ruled out because of the use of a relatively small number of angles, which could have prevented the emergence of non-linear effects (Bruyer et al., 1993; Valentine & Bruce, 1988). None of the recognition and matching studies with various angles has varied the component and configural information independently and data analysis emphasized the role of RTs. To our knowledge, the present study is the first to investigate the effects of orientation upon the detection of component and configural changes using separate analyses of error scores and RTs for several angles of rotation between upright and upside down.

2.3 Experiment 1-2

2.3.1 Method

2.3.1.1 Participants

Sixty-four students from the University of Zurich volunteered as participants in this study. The participants were randomly assigned to one of two groups. In Experiment 1, 16 males and 16 females served as participants. The second group (16 male and 16 female participants) was tested in Experiment 2. All had normal or corrected to normal vision. They were all naïve as to the purpose of this study.

2.3.1.2 Materials

Stimuli were created from grayscale photographs of six people (3 males and 3 females) who had agreed to be photographed and to have their pictures used in psychology experiments. The original grayscale pictures were front facing and with a neutral expression. Digital images were obtained by developing the photographs on Kodak Photo CD™. These images were altered using image-processing software (Adobe Photoshop 4.0 and Canvas 3.5). First, all images were scaled proportionally to have the same interpupilary distance. Then, the hair was removed and the pictures were placed on a black background. These images

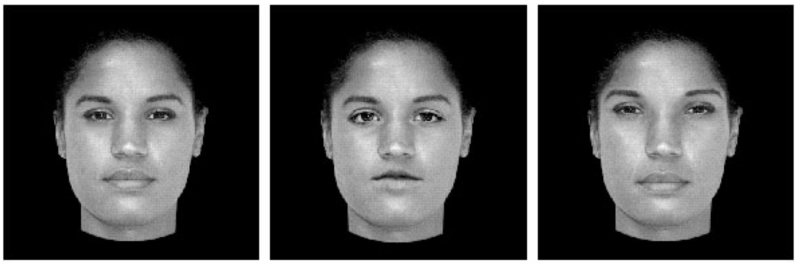

Figure 1. Examples of the stimuli. Left: original face, middle: component change, right: configural change.

constituted the set of six original images. Three anchor points for components were determined: the center of each pupil and the middle of the upper lip contour. The set of six faces with altered component information was created by replacing the eyes and the mouth with components from another face of the same size. The location of new components was the same as in the original images (with an accuracy of one pixel concerning the anchor points defined above). New anchor points were determined in order to produce configural changes. The interpupilary distance, the distance between the pupils and the lower contour of the nose, and the distance between the nose and the mouth were scaled by constant factors (1.16, 1.14, and 1.23, respectively). The eyes and the mouth of the original images were then moved to the new anchor points. This resulted in empty skin areas that were filled with skin patches of the original images in order to ensure a selective change of configural information. All items were copied at seven different orientations (0°, 30°, 60°, 90°, 120°, 150°, 180°). Figure 1 displays examples of the stimuli.

2.3.1.3 Procedure

The experiments were conducted in a dimly lit room. Participants were seated in front of a computer monitor (17 in. screen) at a distance of 1.6 ft (0.48 m). The stimuli covered 10° of visual angle and the viewing distance was maintained by a head rest. A sequential same-different matching task was used. A warning tone (one beep) started each trial. After 300 ms, an upright face was presented for a duration of 3000 ms followed by a 1000 ms blank field. A warning tone (two beeps) announced the second face, which appeared after 300 ms in any one of seven clockwise rotated orientations 0° (upright), 30°, 60°, 90° (horizontal), 120°, 150°, 180° (upside down). The participant indicated whether the two faces were same or different by pressing a key. The participants were instructed to respond as quickly and accurately as possible. Half the participants pressed the same button with their preferred hand and the others used the non-preferred hand. In Experiment 1, "different trials" consisted of faces with altered components (eyes and mouth). In Experiment 2, "different trials" involved faces in which the configural information had been altered. Following the participant's response, a 1000 ms blank field was displayed and the next trial started. Eight random orders were generated using the following constraints: (1) the same orientation was not repeated on consecutive trials, (2) the same face stimulus was not repeated on consecutive trials, and (3) there were no more than four consecutive "same trials" or "different trials". The eight random orders were counterbalanced across the two experiments (component changes vs. configural changes), the gender of the participants and the assignment of the response buttons. There were 84 trials per experiment: 2 (same/different) x 6 (items) x 7 (orientations).

Prior to the experiments proper, a learning session was conducted. First, eight practice trials were performed in order to familiarize the participants with the task. These stimuli were used in the practice trials only. Second, the six experimental pairs consisting of the original and the altered version were shown for five seconds each and the participant was instructed to memorize these pairs. The participants were not informed whether these pairs depicted faces of two different individuals or whether faces of the same individual had been manipulated. The purpose of this learning phase was to allow participants to form upright memory representations of the faces used in the experiment and thereby making the encoding conditions more similar to real-life situations. Third, twelve practice trials were performed (6 "same trials" and 6 "different trials") that contained the experimental face pairs presented sequentially in the upright orientation only. If the participant produced more than one error, these practice trials were repeated once (this occurred for five of the 64 participants).

2.3.2 Results

Individual data were averaged across different faces in order to eliminate an item-specific factor. Separate and combined analyses were carried out on error scores of "different trials" and "same trials". Data were discarded if participants did not respond within 5 seconds. This occurred only in 0.13 per cent of the trials (7 of the 5376 cases).

2.3.2.1 Analysis of error scores

Error scores of "different trials". A two factor analysis of variance (ANOVA) with experiment type (processing of configural vs. component changes) as between-subjects factor and orientation as within-subjects factor was carried out[1] on error scores of "different trials". There were reliable ($p < .001$) main effects of experiment type, $F(1, 62) = 30.53$, $MSE = 0.011$, orientation $F(5, 307) = 15.60$, $MSE = 0.024$, and there was an interaction between experiment type and orientation $F(5, 307) = 11.03$. As it is depicted in Figure 2, changes of orientation had a detrimental effect upon the detection of configural manipulations, whereas the detection of component alterations was much less affected by orientation. In fact, separate one factor within-subjects ANOVAs revealed that the effect of orientation upon the detection of component changes did not reach statistical significance, $F(4, 134) = 1.32$, $MSE = 0.016$, while there was a strong main effect of orientation upon the detection of configural alterations $F(4, 137) = 17.01$, $MSE = 0.050$, $p < .001$.

Since there was no main effect of orientation for the detection of component changes the function relating orientation and error scores was not further analyzed. In contrast, there was a strong effect of orientation upon detecting configural changes, which was further investigated by separate polynomial contrast analyses. There was a significant linear component that was responsible for 40.2 per cent of the variance, $F(1, 31) = 20.87$, $MSE = 0.044$, $p < .001$, a quadratic component $F(1, 31) = 57.69$, $MSE = 0.031$, $p < .001$, explaining 65 per cent of the variance, a cubic component $F(1, 31) = 8.48$, $MSE = 0.032$, p

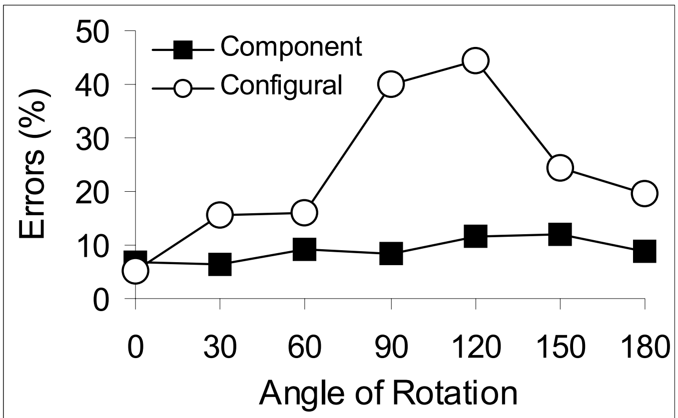

Figure 2. Mean error scores (in per cent) of "different trials" from Experiment 1 (processing of component changes) and Experiment 2 (processing of configural changes).

[1] In all analyses of this study, if Mauchly's (1940) test of sphericity showed a significant deviance (α=0.05) from equicorrelation for a repeated factor or for a combination of factors including at least one repeated factor, Greenhouse and Geisser's (1959) Epsilon was used to adjust the degrees of freedom for the averaged tests of significance.

< .01, accounting for 21.5 per cent of the variance, and an order 5 component $F(1, 31) = 21.10$, $MSE = 0.026$, $p < .001$, responsible for 40.5 per cent of the variance. No other components were significant.

Error scores of "same trials". A two factor ANOVA with experiment type (processing of component vs. configural changes) as between-subjects factor and orientation as within-subjects factor revealed a main effect of orientation $F(4, 261) = 24.78$, $MSE = 0.033$, $p < .001$. There was no effect of experiment type $F(1, 62) = 1.52$, $MSE = 0.019$ but there was an interaction between experiment type and orientation $F(4, 261) = 2.46$, $p < .05$. Separate one factor within-subjects ANOVAs showed a main effect of orientation for the component experiment $F(4, 115) = 18.59$, $MSE = 0.043$, $p < .001$, as well as for the configural experiment $F(4, 136) = 7.00$, $MSE = 0.027$, $p < .001$. As it is depicted in Figure 3, the error scores of "same trials" increased with increasing orientation from upright. This increase was even more pronounced for the "same trials" of the component experiment thus yielding the significant interaction between experiment type and orientation. Separate polynomial contrast analyses on error scores of "same trials" revealed a significant linear component for both experiment types: for the component experiment $F(1, 31) = 47.20$, $MSE = 0.062$, $p < .001$, accounting for 60.4 per cent of the variance, and for the configural experiment $F(1, 31) = 20.41$, $MSE = 0.040$, $p < .001$, responsible for 39.7 per cent of the variance. No other components were significant.

Combined analyses of error scores. Combined analyses of error scores were also carried out in order to investigate the effects of trial type. In the component experiment, the participants made more errors in "same trials" than in "different trials" and this difference increased with increasing orientation from upright. Specifically, a two factor ANOVA with orientation and trial type ("different trials" vs. "same trials") as within-subjects factors showed a main effect of trial type $F(1, 31) = 10.90$, $MSE = 0.075$, $p < .01$, and an interaction between orientation and trial type $F(4, 131) = 9.50$, $MSE = 0.030$, $p < .001$. The same pattern was found in the configural experiment. A two factor ANOVA on error scores of "different trials" and "same trials" revealed also a main effect of trial type $F(1, 31) = 11.28$, $MSE = 0.106$, $p < .01$, as well as an interaction between trial type and orientation $F(4, 137) = 7.65$, $MSE = 0.045$, $p < .001$.

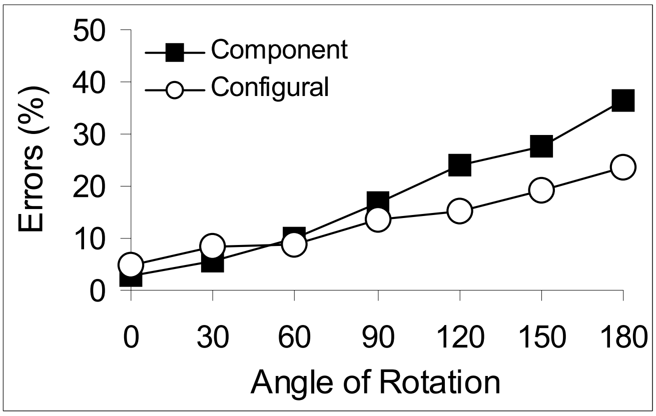

Figure 3. Mean error scores (in per cent) of "same trials" from Experiment 1 (processing of component changes) and Experiment 2 (processing of configural changes).

2.3.2.2 Analysis of RTs

RTs of "different trials". A two factor ANOVA with experiment type (processing of configural vs. component changes) as between-subjects factor and orientation as within-subjects factor on correct RTs of "different trials" revealed a main effect of orientation $F(4, 224) = 18.64$, $MSE = 79{,}140$, $p < .001$. In contrast to the analysis of error scores, the analysis of RTs gave no main effect of experiment type, $F(1, 59) = 0.19$, $MSE = 99{,}993$ and the interaction between experiment type and orientation was not significant, $F(4, 224) = 1.61$. Separate one factor ANOVAs on correct RTs revealed a main effect of orientation for the detection of component changes $F(4, 138) = 12.87$, $MSE = 36{,}645$, $p < .001$, as well as for the detection of configural alterations $F(3, 89) = 8.61$, $MSE = 142{,}380$, $p < .001$. Linear and nonlinear effects in the data displayed in Figure 4 were tested by separate polynomial contrast analyses on correct RTs of "different trials". The linear component was significant in both experiments; for the detection of component alterations, $F(1, 31) = 47.22$, $MSE = 40{,}659$, $p < .001$, accounting for 60.4 per cent of the variance, and for the detection of configural changes, $F(1, 28) = 44.19$, $MSE = 56{,}813$, $p < .001$ responsible for 61.2 per cent of the variance. In both experiments, there was also an order 4 component, which explained 12.3 per cent of the variance for the detection of component changes, $F(1, 31) = 4.34$, $MSE = 28{,}921$, $p < .05$, and 23.2 per cent of the variance for the detection of configural alterations, $F(1, 28) = 8.48$, $MSE = 33{,}227$, $p < .01$. For the processing of configural alterations there was also a quadratic component, $F(1, 28) = 7.74$, $MSE = 63{,}655$, $p < .05$, accounting for 21.7 per cent of the variance. No other components were significant.

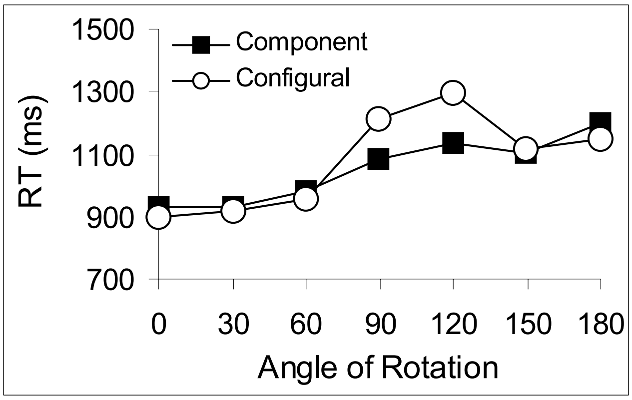

Figure 4. Mean correct RTs (in milliseconds) of "different trials" from Experiment 1 (processing of component changes) and Experiment 2 (processing of configural changes).

RTs of "same trials". A two factor ANOVA on correct RTs of "same trials" revealed a main effect of orientation, $F(4, 248) = 39.40$, $MSE = 68{,}243$, $p < .001$. Like for the error scores, there was no main effect of experiment type for RTs, $F(1, 60) = 2.33$, $MSE = 161{,}323$. The interaction between experiment type and orientation was significant, $F(4, 248) = 3.10$, $p < .05$. Separate one factor within-subjects ANOVAs revealed a main effect of orientation for the component and the configural experiment $F(4, 117) = 26.16$, $MSE = 87{,}354$, $p < .001$, and $F(4, 108) = 15.27$, $MSE = 62{,}572$, $p < .001$, respectively.

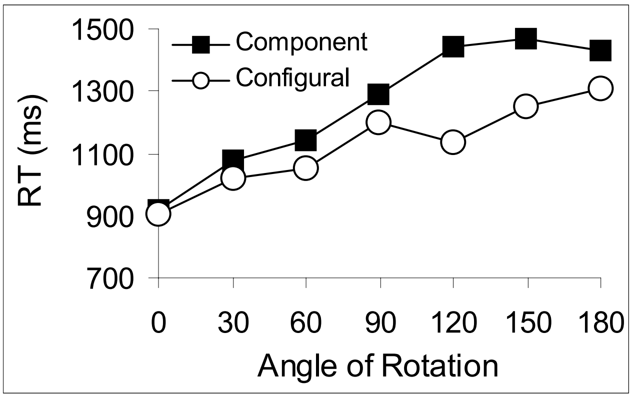

Figure 5. Mean correct RTs (in milliseconds) of "Same trials" from Experiment 1 (processing of component changes) and Experiment 2 (processing of configural changes).

Separate polynomial contrast analyses were also carried out. There was a significant linear effect in the component and the configural experiment, $F(1, 31) = 70.26$, $MSE = 111{,}211$, $p < .001$, explaining 69.4 per cent of the variance, and $F(1, 29) = 42.01$, $MSE = 79{,}615$, $p < .001$, responsible for 59.2 per cent of the variance, respectively. For the configural experiment, there was also an order 6 component, $F(1, 29) = 4.26$, $MSE = 36{,}874$, $p < .05$, which was responsible for 12.8 per cent of the variance. The RT data from the component experiment contained also a quadratic component $F(1, 31) = 11.27$, $MSE = 45{,}475$, $p < .01$, explaining 26.7 per cent of the variance, as well as a cubic component, $F(1, 31) = 4.94$, $MSE = 36{,}404$, $p < .05$, which accounted for 13.7 per cent of the variance. No other components were significant.

Combined analyses of RTs. A two factor ANOVA with orientation and trial type ("different trials" vs. "same trials") as within-subjects factors was carried out on correct RTs of the component experiment. As for error scores, there was a main effect of trial type $F(1, 31) = 33.62$, $MSE = 132{,}450$, $p < .001$, and there was also an interaction between orientation and trial type $F(4, 125) = 6.40$, $MSE = 56{,}394$, $p < .001$.

The same analysis was carried out on correct RTs from the configural experiment. There was no main effect of trial type, $F(1, 26) = 1.14$, $MSE = 114{,}236$, and no interaction between trial type and orientation, $F(3, 85) = 2.13$, $MSE = 106{,}190$.

2.3.3 Discussion

The results from "different trials" indicate that error scores do not depend strongly on orientation when component changes have to be detected, while the participants made up to 44 per cent errors when configural alterations had to be detected. The MDF hypothesis and the schema hypothesis could not predict this result, because neither of them specifies what type of facial information is processed in an orientation sensitive manner. Moreover, this result challenges a purely holistic view of face processing, which implies a disruption of configural and component processing as well. The error scores obtained in Experiment 1 and 2 support the component-configural and the mental rotation hypothesis. Both accounts predict strong impairments by rotation for the detection of configural alterations and small or no effects for the detection of component changes. However, as shown in Table 1, these predictions diverge with regard to RTs. The component-configural hypothesis predicts strong orientation effects for the processing of configural alterations while a separate system for processing component information should not be affected by changes of orientation. This prediction was not supported by the RT data from Experiment 1 and 2. First, the analyses of RTs revealed a main effect of orientation for the detection of configural changes and for the detection of component changes as well. Second, there was no main effect of experiment type, nor was there an interaction between orientation and experiment type. And third, as revealed by polynomial contrast analyses, the increase of RTs with increasing orientation was described best by a linear trend, which accounted for the processing of configural information as well as for the processing of component information.

While these results do not confirm with the predicted effects by the component-configural hypothesis they are highly compatible with the predictions derived from the mental rotation hypothesis. According to this theoretical view, complex stimuli such as faces overtax a mental rotation mechanism when they are substantially disoriented. Consequently, facial parts need to be mentally rotated one after the other and RTs generally increase with increasing orientation. The rotation of facial parts or components can not be done simultaneously and therefore, information about the spatial or configural relationships is much less recoverable (Valentine & Bruce, 1988). This prediction is consistent with the accuracy data obtained in experiment 1 and 2, which revealed high error scores for the detection of configural alterations in disoriented faces while such a profound orientation sensitivity was absent for the detection of component changes.

Further evidence in favor of the mental rotation hypothesis is provided by a comparison between "different trials" and "same trials" in the component experiment (Experiment 1). The mental rotation hypothesis predicts that in "same trials" more components have to be mentally rotated in order to check for a potential mismatch as opposed to "different trials", which actually contained a difference. On average, participants could stop mentally rotating components as soon as they detected one change and therefore, they had to process less facial features in the "different trials" than in the ”same trials” of the component experiment. And indeed, participants needed more time and made more errors on "same trials" when compared to "different trials" and this difference increased with increases in rotation from upright.

It is noteworthy that this prediction does not account for detecting configural changes (Experiment 2) since all faces contained identical components and differed only in configural information. Hence, there appears no reason to expect large differences between RTs of "different trials" and "same trials". And indeed, the results showed neither an effect of trial type nor an interaction between trial type and orientation for the RTs of the configural experiment.

Finally, it should be noted that the error scores from "different trials" in the configural experiment did not increase monotonically from upright to inverted. Instead, most errors occurred when the stimuli were presented at intermediate orientations of 90° and 120° (see Figure 2). We reserve the discussion of this effect for the General Discussion.

The general pattern of results from experiments 1-2 provided a large body of evidence in favor of the mental rotation hypothesis, whereas other theories failed to predict the differential effects of orientation. It should be pointed out that the results of the current study do not allow decisive conclusions regarding the representation of *upright* faces, more specifically whether upright faces are represented only holistically or by component and configural representations. However, our results clearly indicate that the mental rotation hypothesis provides a predictive utility for the processing of *disoriented* faces. In fact, proponents of holistic as well as of component and configural representations have noted the explanatory power of the mental rotation hypothesis. For example, Farah et al. (1995) have pointed out that the deeper answer to the question "Why is face recognition so orientation sensitive? ... will concern capacity limitations of the orientation normalization process." (Farah et al. 1995, p.633). Similarly, Searcy and Bartlett (1996) mentioned that the difficulty of processing configural information in disoriented faces could be due to capacity limitations of a mental rotation mechanism. To our point of view, the critical question is whether *rotated* faces are processed by *two separate mechanisms* for component and configural information or whether there are two separate types of memory representations (componential and configural/holistic) that are accessed by *one mechanism* (mental rotation).

The investigation of transfer effects is a promising approach in further revealing the underlying information processing structure (e.g., Fahle, 1997). A transfer effect exists when learning experience acquired in one task (e.g., the detection of component alterations) increases performance in a different task, which is tested subsequently (e.g., the detection of configural changes). The occurrence of transfer effects is normally produced by mechanisms that are shared by both tasks. Presumably, mental rotation is a common mechanism in the experiments of the present study.

2.4 Experiments 3-4

A transfer effect could be expected if mental rotation were needed for the detection of component and configural alterations in rotated faces. In contrast, transfer effects should be absent if detecting component and configural changes would involve two separate processing systems, as it is proposed by a strict dual-mode interpretation of the component-configural hypothesis. The participants that were tested in the processing of component information (Experiment 1) were now tested in the processing of configural information (Experiment 4). The other half of the participants had performed the configural experiment first (Experiment 2) and were subsequently tested in the processing of component changes (Experiment 3). Thus, this design allowed testing for transfer effects by comparing performance in the same experimental task, which was either conducted first or second.

2.4.1 Method

The participants, materials, apparatus, design, and procedures were identical to experiments 1-2. The only difference was a reversal of the experimental task (detection of component vs. configural changes), which was conducted following a short break after the first experiment.

2.4.2 Results

Before being subjected to analyses, the data of each participant were averaged across the different faces in order to eliminate an item-specific factor. Data were discarded if participants did not respond within 5 seconds after the exposure of the test face. This occurred for 0.11 per cent of all trials (6 of the 5376 cases).

2.4.2.1 Analysis of error scores

Separate analyses of transfer effects in error scores of ”different trials” and "same trials" were carried out.

"Different trials". Separate two factor ANOVAs with presentation condition (experiment conducted in the first vs. second block) as between-subjects factor and orientation as within-subjects factor were carried out on error scores of "different trials". There was no transfer effect for the detection of component changes, since there was no effect of presentation condition, $F(1, 62) = 2.04$, $MSE = 0.035$, and there was no interaction between orientation and presentation condition, $F(5, 296) = 0.73$, $MSE = 0.012$. For the detection of configural alterations, there was a transfer effect as indicated by the main effect of presentation condition $F(1, 62) = 10.22$, $MSE = 0.126$, $p < .01$. This transfer effect was dependent on orientation as indicated by the interaction between orientation and presentation condition $F(5, 293) = 2.66$, $MSE = 0.037$, $p < .05$.

"Same trials". There was a transfer effect for both task types as indicated by the main effects of presentation condition, $F(1, 62) = 5.64$, $MSE = 0.099$, $p < .05$, and $F(1, 62) = 4.90$, $MSE = 0.081$, $p < .05$, respectively. For both task types, the effects of transfer were not dependent on orientation since there was no interaction between orientation

and presentation condition, $F(4, 217) = 2.34$, $MSE = 0.037$, and $F(5, 281) = 0.33$, $MSE = 0.023$, respectively.

2.4.2.2 Analysis of reaction times

The same procedures were used for the analyses of transfer effects in correct RTs of "different trials" and "same trials".

"Different trials". Separate two factor ANOVAs were carried out on correct RTs of "different trials" from the two tasks. There was no transfer effect for the detection of component changes, since there was no effect of presentation condition $F(1, 62) = 0.41$, $MSE = 562{,}069$, and there was no interaction between orientation and presentation condition, $F(5, 297) = 0.79$, $MSE = 32{,}223$. For the detection of configural alterations, there was a transfer effect as indicated by the main effect of presentation condition $F(1, 59) = 4.26$, $MSE = 720{,}083$, $p < .05$. This effect was independent of orientation since there was no interaction between orientation and presentation condition $F(4, 227) = 1.40$, $MSE = 102{,}631$.

"Same trials". There were no indications of transfer effects for the correct RTs of "same trials". The main effect of presentation condition was not significant neither for the component experiments, $F(1, 62) = 2.15$, $MSE = 893{,}250$, nor for the configural experiments, $F(1, 60) = 2.65$, $MSE = 752{,}915$. Moreover, there was no interaction between orientation and presentation condition, neither for the component, nor for the configural task, $F(4, 251) = 0.95$, $MSE = 78{,}509$, and $F(4, 259) = 1.17$, $MSE = 49{,}760$, respectively.

2.4.3 Discussion

In Experiments 3-4, we investigated whether the processing of disoriented faces can be explained on the basis of one single mechanism such as mental rotation, or whether two separate mechanisms are needed to account for the results (e.g., two different modes for the processing component and configural information). This was achieved by testing for transfer effects. A transfer effect is apparent if in one task (e.g., the detection of component alterations) a learning effect occurs, which results in a better performance on a subsequent task (e.g., the detection of configural alterations). It follows from the existence of a transfer effect, that both tasks must share a common mechanism.

A comparison of the accuracy data from experiments 1-4 revealed several transfer effects. Performing one task usually reduced the error scores in the second task. This result is difficult to reconcile with the idea of two separate and independent mechanisms for the processing of component and configural information. In contrast, the finding of transfer effects is consistent with the assumption of a mental rotation mechanism that is needed in both experiments and becomes more accurate through practice. Noticeably, the speed of mental rotation remained relatively stable from one experiment to the other as indicated by the absence of transfer effects for RT’s in three of four possible cases.

2.5 General discussion

The analyses of error scores from experiments 1-2 revealed that orientation had a non-significant effect upon the detection of component changes while the detection of configural alterations was strongly impaired when faces were substantially rotated away from the upright position. Neither the MDF hypothesis nor the schema hypothesis could predict this result since none of these accounts specifies selective orientation effects upon the processing of these two types of facial information. Moreover, this result poses problems for a purely holistic view of face processing, which implies that rotating a face disrupts the processing of what is nominally component and configural information. At the same time, the error scores obtained in the experiments supported the component-configural hypothesis as well as the mental rotation hypothesis. They both predict strong impairments by rotation for the detection of configural alterations while the detection of component changes should be affected much less if at all. However, these two accounts for the face inversion effect differ in their predictions on RTs. According to the mental rotation hypothesis faces are so complex that they overtax an orientation normalization mechanism, so that the subjects would try to detect component and configural changes by mentally rotating faces part by part. This would lead to an increase of RTs with increasing rotation from upright in both experiments. This prediction was confirmed in this study. The RTs for detecting component and configural changes increased with increasing angular disparity following a similar linear trend. Incompatible with this result, the component configural hypothesis posits that rotating faces has a strong effect upon a mechanism for processing configural information, while a separate system for processing of component changes should not be substantially affected by changes of orientation. As second step, we carried out an analysis of transfer effects. If faces have to be mentally rotated part by part, mental rotation is needed for the detection of component as well as configural alterations. By comparing the results from Experiment 1 to Experiment 3 and Experiment 2 to Experiment 4 several transfer effects have been revealed from one experimental task to the other. This result is compatible with the assumption that mental rotation is needed to access orientation-bound representations in order to detect component as well as configural alterations. At the same time, these transfer effects are difficult to reconcile with the idea of two separate and independent modes for processing component and configural information. In short, the findings from experiment 1-4 are consistent with the mental rotation hypothesis and no other account yields a comparatively coherent explanation for the pattern of results observed in this study.

However, there was a somewhat unexpected finding for the error scores of detecting configural changes. Instead of a monotonic increase, "different trials" of Experiment 2 and 4 showed that participants made most errors at intermediate orientations of 90° and 120° and not when the faces were presented upside-down. Remarkably, a similar effect has been found in object naming studies**.** The time to name line drawings of natural objects has been found to increase linearly from upright to 120° of planar rotation, while naming times for 180° are often faster than those for 120° (e.g., Jolicoeur, 1985; Murray, 1995a, 1995b, 1997). However, such nonlinear effects are present primarily on the initial trials only; after practice, they are usually diminished or even disappear. In fact, recent studies suggest that when the stimulus set contains orientation-invariant information, effects of orientation disappear following experience (Murray, 1999), which can occur even after a single presentation of objects in a block of trials (Murray, Jolicoeur, McMullen, & Ingleton, 1993). Interestingly, in our study strong effects of orientation remained stable even after a remarkable amount of practice. This is consistent with the view that a transition to orientation-invariant processing could not take place and the subjects had to rely on normalization mechanisms for detecting component and configural alterations. An explanation for nonlinear effects of orientation has been provided by Corballis, Zbrodoff, Shetzer, and Butler (1978). They suggested that it might be possible to "mentally flip" an inverted picture out of the plane to match it to a memory representation (see also Koriat, Norman, & Kimchi, 1991). Based on this idea, one could assume that for rotation angles higher than 90° or 120° facial features first mentally rotated upside down then immediately flipped to upright in order to match them to orientation-bound upright memory representations. If mental flipping is a faster normalization procedure, such an interpretation would predict higher RTs (and error scores) for intermediate as opposed to upside-down orientations. This could explain why in the present study substantial nonlinear effects were found in the error scores and (less pronounced) in the RT curves of "different trials" of the configural experiments.

It should be noted however, that this nonlinear effect was much less apparent in "same trials" of the configural experiments and for both "same" and "different trials" when component changes had to be detected. Why did the participants not mentally flip in those trials? Therefore, an alternative explanation for this nonlinear effect is by all means conceivable.

The nonlinear effects found for the detection of configural alterations could also be related to the phenomenon known as horizontal-vertical illusion (HVI); a vertically oriented line appears longer than a horizontally oriented line of exactly the same physical length. The HVI was first reported by Fick (1851) and it affects the perception of various objects including complex stimuli such as houses (e.g., Higashiyama, 1996; Yang, Dixon, & Proffitt, 1999). Künnapas (1955) used a T shaped stimulus and showed that an upright T (T) or an inverted T (⊥) produced a greater illusion than a 90° rotated T (-| or |-). Similar results have been revealed already by Finger and Spelt (1947) and subsequently by Tedford and Tudor (1969). In the present study, configural alterations in faces were accomplished by increasing the distance between the eyes and between the eyes and mouth. As a matter of fact, the distances between these features form a T shape as illustrated in Figure 6. Even though all lines are equally long the dividing line in the picture on the left appears to be substantially longer than in the picture on the right. If the same mechanisms underlying the HVI are also involved in encoding faces, the distance between the eyes and the mouth would be overestimated in an upright face. However, when the face is rotated 90°, the distance between the eyes and the mouth is horizontal and therefore appears shorter. Consequently, it could be assumed that it is much more difficult to correctly detect configural alterations when a face is rotated 90°. The increased distance between eye and mouth appears shorter at 90° and therefore it is possible that a face with increased configural information appears to better match with the unchanged upright face. The error rate would decrease again as soon as the face is

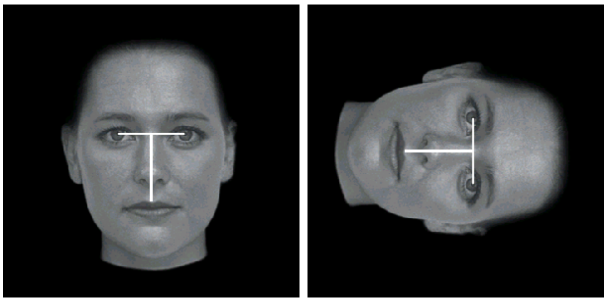

Figure 6. Horizontal-vertical illusion and perception of configural information in faces. Although all lines are of the same geometrical length, the dividing line in the picture on the left is perceived as being substantially longer than in the picture on the right.

rotated beyond 90°. Therefore, such an effect could have caused the quadratic trend component in the error scores of "different trials" when configural changes had to be detected. Note that such an interpretation only holds if the subjects based their decisions dominantly upon the perceived eye-mouth distance. If the responses would have been based on the perceived inter-eye distance, exactly the opposite predictions would follow from the HVI hypothesis (i.e. lowest error scores for intermediate orientations in “different trials” of the configural experiments). The answer to the question whether the nonlinear effect found in this study can be explained by flipping, by the HVI or by combinations of both is beyond the scope of the present article and is currently addressed in separate experiments (Schwaninger & Ryf, in preparation).

In returning to the overall picture, several conclusions with regard to the processing of faces can be drawn from this study. First, the finding that component changes could be detected relatively independent of orientation while detecting configural alterations was detrimentally impaired in rotated faces clearly indicates the existence of explicit featural representations, whether they bear a hierarchical relation to whole face representations, or whether they constitute an independent population of representations.

Second, as a consequence, a purely holistic view of face processing (i.e. no explicit part-based representations), is not compatible with the results of the present study. It should be stressed however that for the processing of *upright* faces, holistic representations could play a pivotal role, which would be consistent with our finding that participants were able to detect configural changes with the same accuracy as component alterations when the faces were presented upright. But, as evidently shown, when faces are substantially rotated away from upright, matching face representations as unparsed perceptual wholes is strongly impaired because rotated faces seem to overtax a mental rotation mechanism and the faces have to be processed by mentally rotating facial parts. Neurophysiological evidence supports the assumption of part-based as well as configural and holistic face representations. For some neurons, selectivity for particular features of the head and face, for example the eyes and mouth has been revealed (Perret, Rolls, & Caan, 1982; Perret, Mistlin, & Chitty, 1987; Perret, Hietanen, Oram, & Benson, 1992). Other groups of cells need the simultaneous presentation of multiple parts of a face and are therefore related to a more holistic type of representation (Perret & Oram, 1993; Wachsmuth, Oram, & Perret, 1994). Yamane, Kaji, & Kawano, (1988) have discovered neurons that seem to process configural information since they detect combinations of distances between facial parts, such as the eyes, mouth, eyebrows, and hair.

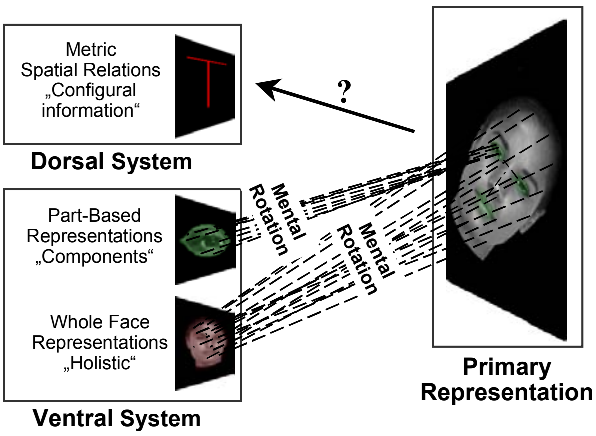

Figure 7. Integrative model of face recognition.

A third conclusion that can be drawn from this study is that face recognition relies on separate representations for component and configural information. The transfer effects from the processing of component alterations to the processing of configural changes and vice versa are absolutely consistent with the assumption that one mechanism like mental rotation is needed for accessing part-based representations and configural/holistic representations. Although it would be a simple matter, computationally, to rotate a pictorial face representation, the large cost to human recognition performance and the results from the present study indicate that mentally rotating a face as a whole overtaxes the capacity limitations of such an orientation normalization process. These conclusions are summarized in Figure 7, which depicts an integrative model of face processing. Initially, visual stimuli are assumed to be represented as a pictorial and metric representation in primary visual areas. The recognition of a face requires the extraction of enough information from the primary representation in order activate a memory representation in higher visual areas. Such a representation can be envisaged as a distributed network of detectors for relatively orientation-invariant information like color and texture, moderately orientation-sensitive information like facial components, and detectors for highly orientation-bound information, which entail holistic and/or configural aspects of a face. Depending on the recognition task, the system will be more or less sensitive to orientation. If faces have to be distinguished from houses, orientation-invariant properties like color and texture would suffice and performance should be relatively independent of orientation (e.g., Valentine & Bruce, 1988). In contrast, however, if an individual has to be recognized, the task requires the detection of subtle featural and configural differences which relies on extensive expertise with upright faces (Diamond & Carey, 1986; Carey, 1992). Because faces are usually seen upright, orientation-sensitive memory representations need to be accessed through normalization mechanisms (e.g. mental rotation and flipping). Rotating a face as a whole overtaxes such mechanisms which causes substantial impairment in activating orientation-bound part-based and in particular configural/holistic face representations. This could be the deeper answer to the question ”Why is face recognition so orientation sensitive”.

2.6 References

Bartlett, J.C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281-316.

Bruce V. (1988). *Recognizing faces.* Hillsdale, NJ: Erlbaum.

Bruyer, R., Galvez, C., & Prairial, C. (1993). Effect of disorientation on visual analysis, familiarity decision and semantic decision on faces. *British Journal of Psychology, 84,* 433-441.

Carey S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London, B, 335,* 95-103.

Carey, S., & Diamond, R. (1977). From piecemal to configurational representation of faces. *Science, 195,* 312-314.

Cochran, E. L., Pick, A.D., & Pick, H.L. (1983). Task-specific strategies of mental "rotation" of facial representations. *Memory & Cognition, 11,* 41-48.

Corballis, M.C., Zbrodoff, N.J., Shetzer, L.I., & Butler, P.B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6,* 98-107.

Diamond, R., & Carey, S. (1986). Why faces are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Ellis, H.D. (1975). Recognizing faces. *British Journal of Psychology, 66,* 409-426.

Ellis, H.D. (1992). The development of face processing skills. *Philosophical Transactions of the Royal Society of London, B, 335,* 105-111.
Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Research, 37,* 1885-1895.
Farah, M.J., Drain, H.M., & Tanaka, J.W. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628-634.
Farah, M.J., Wilson, K.D., Tanaka, J.N., & Drain, M. (1998). What is ”special” about face perception? *Psychological Review, 105,* 482-498.
Fick, A. (1851). *De errore quodam optico asymenetria bulbi effecto.* Marburg: J.A. Kochii.
Finger, F.W., & Spelt, D.K. (1947). The illustration of the horizontal-vertical illusion. *Journal of Experimental Psychology, 37,* 243-250.
Goldstein, A.G., & Chance, J.E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105,* 47-59.
Goldstein, A.G., & Chance, J.E. (1981). Laboratory studies of face recognition. In G.M. Davies, H.D. Ellis, & J.W. Sheperd (Eds.), *Perceiving and remembering faces.* London, England: Academic Press.
Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 32,* 95-112.
Higashiyama, A. (1996). Horizontal and vertical distance perception: The discorded-orientation theory. *Perception & Psychophysics, 58,* 259-270.
Johnston, R.A., & Ellis, H.D. (1995). The development of face recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition, explorations in face space* (pp. 1-23). London, England: Routledge.
Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289-303.
Köhler, W. (1940). *Dynamics in Psychology.* New York: Liveright.
Koriat, A., Norman, J., & Kimchi, R. (1991). Recognition of rotated letters : Extracting invariance across successive and simultaneous stimuli. *Journal of Experimental Psychology: Human Perception & Performance, 17,* 444-457.
Künnapas, T.M. (1955). An analysis of the "vertical-horizontal illusion”. *Journal of Experimental Psychology, 49,* 134-140.
Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology, A, 5,* 449-73.
Mauchly, J.W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics, 11,* 204-210.
Murray, J.E. (1995a). Imagining and naming rotated natural objects. *Psychonomic Bulletin & Review, 2,* 239-243.
Murray, J.E. (1995b). Negative priming by rotated objects. *Psychonomic Bulleting & Review, 2,* 534-537.
Murray, J.E. (1997). Flipping and spinning: Spatial transformation procedures in the identification of rotated natural objects. *Memory & Cognition, 25,* 96-105.
Murray, J.E. (1999). Orientation-specific effects in picture matching and naming. *Memory & Cognition, 27,* 878-889.
Murray, J.E., Jolicoeur, P., McMullen, P.A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition, 21,* 604-610.
Murray, J.E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11,* 492-496.
Perret, D.I., & Oram, M.W. (1993). Neurophysiology of shape processing. *Image and Vision Computing, 11,* 317-333.

Perret, D.I., Hietanen, J.K., Oram, M.W., & Benson, P.J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, B, 335,* 23-30.

Perret, D.I., Mistlin, A.J., & Chitty, A.J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience, 10,* 358-364.

Perret, D.I., Rolls, E.T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329-342.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition, 47,* 25-57.

Rock, I. (1973). *Orientation and form.* San Diego, CA: Academic Press.

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230,* 78-85.

Rock, I. (1988). On Thompson's inverted-face phenomenon (Research Note). *Perception, 17,* 815-817.

Scapinello, K.F., & Yarmey, A.D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21,* 29-33.

Schwaninger, A., & Ryf, S. (in preparation). Low vs. high level vision: Does the horizontal vertical illusion affect configural face processing?

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904-915.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75,* 221-242.

Sergent, J., & Corballis, M.C. (1989). Categorization of disoriented faces in the cerebral hemispheres of normal and commissurotomized subjects. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 701-710.

Stevenage, S.V. (1995). Expertise and the caricature advantage. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition, explorations in face space* (pp. 24-46). London, England: Routledge.

Tanaka, J.W., & Farah, M.J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50,* 367-372.

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A,* 225-245.

Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25,* 583-592.

Tedford, W.H., & Tudor, L.L. (1969). Tactual and visual illusions in the T-shaped figure. *Journal of Experimental Psychology, 81,* 199-201.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9,* 483-484.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161-204.

Valentine, T. (Ed.). (1995). *Cognitive and computational aspects of face recognition, explorations in face space.* London, England: Routledge.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition, 16,* 556-566.

Wachsmuth, E., Oram, M.W., & Perret, D.I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex, 4,* 509-522.

Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Experimental Brain Research, 73,* 209-214.

Yang, T.L., Dixon, M.W., & Proffitt, D.R. (1999). Seeing big things: Overestimation of heights is greater for real objects than for objects in pictures. *Perception, 28,* 445-467.

Yarmey, A.D. (1971). Recognition memory for familiar ”public” faces: Effects of orientation and delay. *Psychonomic Science, 26,* 286-288.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141-145.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception, 16,* 747-759.

3 Configural information is processed differently in perception and recognition of faces

3.1 Abstract

Several previous studies have stressed the importance of processing configural information in face *recognition*. In this study the *perception* of configural information was investigated. Large overestimations were found when the eye-mouth distance and the inter-eye distance had to be estimated. Whereas configural processing is disrupted when inverted faces have to be recognized the perceptual overestimations persisted when faces were inverted. These results suggest that processing configural information is different in perceptual as opposed to recognition tasks.

3.2 Introduction

Processing facial information is one of the most relevant skills in everyday life. Although faces seem to look quite different from each other, they do in fact form a very homogenous stimulus class when seen from an image-based point of view. Each face has the same components (eyes, nose, mouth etc.) in the same basic arrangement. Therefore, reliably recognizing faces entails detecting subtle differences between components and their spatial interrelationship (configural information). Whereas component processing seems to be relatively unaffected by orientation changes the processing of configural information is strongly impaired when faces are rotated. Indeed, many researchers have argued that turning faces upside-down disrupts configural processing much more than component processing (e.g. Leder & Bruce, 2000; Murray, Yong, & Rhodes, 2000; Schwaninger & Mast, 1999; Searcy & Bartlett, 1996; Sergent, 1984). More than 30 years ago, it was found that face recognition is disproportionately affected by inversion when compared to the recognition of other mono-oriented objects such as airplanes, houses, and stick figures of men in motion (Yin, 1969). Since face recognition is highly orientation-sensitive and the processing of configural information is strongly impaired when faces are turned upside-down many researchers have devoted a special role to processing configural information in face recognition. Whereas many previous studies have investigated the role of configural information for recognizing faces this study examines the perception of configural information in upright and rotated faces.

3.3 Experiment 5

Face recognition is characterized by a high sensitivity for configural information. For example Haig (1984) revealed for unfamiliar faces that configural alterations, which were induced by changing the distance between facial components are sometimes detected at the visual acuity threshold level. Similar results were reported by Hosie, Ellis and Haig (1988) for familiar faces. Whereas these studies were concerned with detecting alterations of configural information in faces the aim of Experiment 5 was to investigate whether human observers have a veridical percept of configural information.

3.3.1 Method

3.3.1.1 Participants

Twenty undergraduates from the University of Zurich voluntarily participated in this study. The participants were randomly assigned to two groups of 10 participants. All had normal or corrected to normal vision.

3.3.1.2 Materials and procedure

Photographs were made from 10 people (5 females) who had agreed to be photographed and to have their pictures used in psychology experiments. The faces in the original grayscale pictures were front facing and had a neutral expression. In digital versions the hair was removed and the faces were placed on a black background. The experiments were conducted in a dimly lit room. The viewable screen area on the TFT display was limited to a 750*750 pixel square (23.5° of visual angle) by a cardboard covering the 14.1 inch screen. The viewing distance was maintained by a head rest so that the center of the screen was at eye height of participants and the height and width of displayed faces covered 8.5° and 6.7° of visual angle, respectively.
The method of adjustment was applied. The length of a simultaneously presented white line (comparison stimulus) had to be adjusted in order to appear as long as the standard stimulus. For half the participants the standard stimulus was the eye-mouth distance, for the other half of participants the standard stimulus was the inter-eye distance (Figure 1). The latter was defined as the distance between the pupils (mean distance was 84 pixel or 2.6° of visual angle). The eye-mouth distance was defined as the vertical distance between the point in the middle of the upper contour of the mouth and the point where a vertical line through this point would cross a horizontal line connecting the two pupils (mean distance was 86 pixel or 2.7° of visual angle). Adjustments were made with the preferred hand by turning a small wheel on a mouse device. Each trial was started by pressing a button on this device. The adjustment line (comparison stimulus) was one pixel in width and its initial length was either 20 or 180 percent of the standard stimulus. For the two standard stimuli (inter-eye distance and eye-mouth distance) the line comparison stimulus was presented horizontally to the right of the standard stimulus and vertically on bottom of the standard stimulus (Figure 8). There were 40 trials for each standard stimulus: 10 (faces) * 2 (initial line lengths) * 2 (positions). The order of faces, initial line lengths, and line positions was counterbalanced across participants using latin squares.

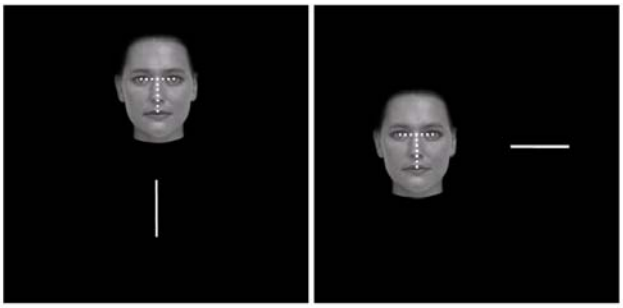

Figure 8 The two positions of standard and comparison stimuli (line) for one face as standard stimulus. Dotted lines indicate the inter-eye distance and eye-mouth distance and were not shown in the experiments.

3.3.2 Results and discussion

Individual data were averaged across the two measurement conditions, the two initial line lengths and the ten faces. The eye-mouth distance was overestimated by 39 percent (*SE* = 5.96) and the inter-eye distance by 11 percent (*SE* = 4.02)[2].

Several previous studies have found a high sensitivity for detecting subtle configural changes (Bruce, Doyle, Dench, & Burton, 1991; Haig, 1984; Hosie et al., 1988; Kemp, McManus, & Pigott, 1990). The large overestimations revealed in the present study indicate that the ability of skilled perceptual discrimination does not necessarily imply very precise veridical percepts. In contrast, the overestimations found in Experiment 5 are of a magnitude that exceeds most known perceptual size illusions (e.g. Coren & Girgus, 1978).

3.4 Experiment 6

The processing of configural information in recognition and detection tasks is strongly impaired when faces are inverted (Leder & Bruce, 2000; Rhodes, Brake, & Atkinson, 1993; Schwaninger & Mast, 1999; Sergent, 1984; Young, Hellawell, & Hay, 1987). If there was a difference in the perception of configural distances between upright and inverted faces, then the face inversion effect could be related to perceptual processes. In contrast, if the overestimations found in Experiment 5 would persist to the same degree in inverted faces, the orientation-dependent nature of configural processing in face recognition can not be explained based on limitations on the perceptual level.

A second aim of Experiment 6 was to investigate a possible role of the horizontal vertical illusion (HVI). This perceptual phenomenon has been first reported by Fick (1851) and refers to the observation that vertical lines or distances appear longer than horizontal ones of the same physical length. The HVI has been shown to affect also the perception of various objects including complex stimuli such as houses (e.g. Higashiyama, 1996; Yang, Dixon, & Proffitt, 1999). In Experiment 6 a potential effect of the HVI upon the perception of configural information in faces was investigated by showing the faces in four angles of clockwise rotation (0°, 90°, 180°, 270°) and comparing the overestimations of configural information to the overestimation of line length.

3.4.1 Method

3.4.1.1 Participants

Twenty-four undergraduates from the University of Zurich volunteered in this study. All had normal or corrected to normal vision.

---

[2] Based on the horizontal vertical illusion (HVI), the horizontal vs. vertical placement of the comparison line could be expected to influence the adjustments. Indeed, separate analyses for the two measurement conditions (horizontal vs. vertical placement of the comparison line) revealed for both facial distances significant effects: When the comparison line was horizontally oriented (as opposed to vertically oriented), the overestimation of the eye-mouth distance was 10 percent larger, $t(9) = 2.98$, $p < .05$, and the overestimation of the inter-eye distance was 8 percent larger, $t(9) = 3.71$, $p < .01$. In order to reduce such effects based on the placement of the comparison line, the data were averaged across the two measurement conditions.

3.4.1.2 Materials and procedure

One male and one female face from Experiment 5 served as stimuli. The experimental setup was identical to Experiment 5. The length of a simultaneously presented white line (comparison stimulus) had to be adjusted in order to appear as long as the standard stimulus. For 12 randomly selected participants the standard stimulus was the inter-eye distance and the eye-mouth distance of the simultaneously presented face (both distances were 83 pixel or 2.6° of visual angle). The distances were explained to the participants the same way as in Experiment 5. In order to ensure that the participants understood the definitions of the distances precisely the distances were indicated with white lines on a face presented on a cardboard above the computer screen. The eye-mouth and the inter-eye distance were adjusted in separate blocks, counterbalanced across subjects. For the other 12 randomly selected participants the standard stimulus was a simultaneously presented white line that was one pixel in width and 83 pixel in length. Adjustments were made as in Experiment 5. Again, the adjustment line (comparison stimulus) was one pixel in width and its length was either 20 or 180 percent of the standard stimulus. The comparison stimulus was presented horizontally to the right or left of the standard stimulus and vertically on top or bottom of the standard stimulus, so that in half the trials the comparison line was at the same orientation as the facial distance, whereas in the other half of the trials the comparison line was perpendicular to it. The standard stimuli were presented in four angles of clockwise rotation (0°, 90°, 180°, 270°) around their center.

There were two blocks of 64 trials resulting in 128 trials for the group in which the eye-mouth distance and the inter-eye distance served as standard stimuli: 2 (adjustments for the male and female face) * 2 (initial lengths of comparison stimulus) * 4 (positions of standard and comparison stimuli) * 4 (angles of rotation of the standard stimulus) * 2 (blocks: eye-mouth distance and inter-eye distance). Since for the second group the standard stimulus was a line instead of facial distances only one block (64 trials) was used: 2 (adjustments) * 2 (initial lengths of comparison stimulus) * 4 (positions of standard and comparison stimuli) * 4 (angles of rotation of the standard stimulus). The order of positions, rotations, length of comparison stimulus as well as order of faces and blocks (group one only) was counterbalanced across participants using a mixed latin square design.

3.4.2 Results and discussion

Individual data were averaged across the four measurement conditions, the two initial lengths of the comparison stimulus as well as the two adjustments. As shown in Figure 9, the line was overestimated when presented vertically and slightly underestimated when presented horizontally. This result reflects the well known horizontal

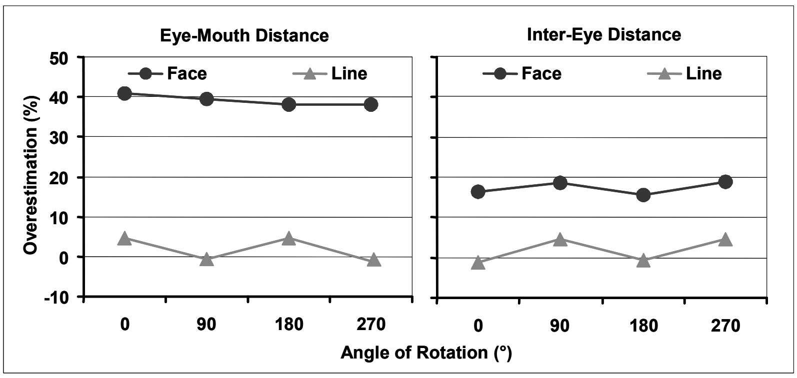

Figure 9 Large overestimation of configural information in faces and the effect of orientation. Left: eye-mouth distance, right: inter-eye distance.

vertical illusion. The results from Experiment 5 were replicated. The eye-mouth distance was overestimated by 41 percent and the inter-eye distance by 16 percent in upright faces[3]. A two factor analysis of variance (ANOVA) with standard stimulus (eye-mouth distance vs. line) as between-subjects factor and orientation as within-subjects factor revealed that the eye-mouth distance was much more overestimated than the line, $F(1, 22) = 13.79$, $MSE = 2422.01$, $p < .01$. There was also a main effect of orientation[4], $F(2.33, 51.23) = 18.89$, $MSE = 6.16$, $p < .001$, and an interaction between orientation and standard stimulus (eye-mouth distance vs. line), $F(2.33, 51.23) = 10.73$, $p < .001$. As indicated by the interaction the HVI affected perceived line length more than the perception of the eye-mouth distance.

		Eye-Mouth Distance			Inter-Eye Distance		
(I) ANGLE	(J) ANGLE	MD (I-J)	*SE*	*p*	MD (I-J)	*SE*	*p*
0	90	1.623	1.306	1.000	-2.170	1.108	.456
0	180	2.843	1.110	.159	0.855	1.117	1.000
0	270	2.930	1.046	.103	-2.496	1.044	.215
90	180	1.220	1.045	1.000	3.025	1.213	.179
90	270	1.306	0.708	.552	-0.326	0.623	1.000
180	270	0.087	0.776	1.000	-3.351	1.139	.080

Table 2 Bonferroni corrected pairwise comparisons between the four angles used in Experiment 2. *Note.* MD = mean difference, *SE* = standard error.

A separate two factor analysis of variance (ANOVA) with standard stimulus (inter-eye distance vs. line) as between-subjects factor and orientation as within-subjects factor revealed larger overestimations of the inter-eye distance than of line length, $F(1, 22) = 4.86$, $MSE = 1177.18$, $p < .05$. There was a main effect of orientation, $F(2.28, 50.09) = 26.90$, $MSE = 6.63$, $p < .001$. Again, there was an interaction between orientation and standard stimulus (inter-eye distance vs. line), $F(2.28, 50.09) = 3.19$, $p < .05$, confirming that also the perception of the inter-eye distance is less affected by the HVI than the perception of lines. The effects of orientation were further examined using Bonferroni corrected pairwise comparisons of means (Table 2). There were no significant differences neither for the inter-eye distance nor for the eye-mouth distance. More specifically, the large overestimations

[3] As mentioned in footnote 1, the placement of the comparison line had a modulatory effect on the overestimations in Experiment 5. Similar effects were found in Experiment 6. On average, the overestimation was 8 percent larger for horizontal vs. vertical placements of the comparison line. This effect was comparable across conditions since separate ANOVAs for the eye-mouth and the inter-eye distance with measurement condition as within-subjects factor (horizontal vs. vertical placement of the comparison line) and standard stimulus (line vs. facial distance) as between-subjects factor gave no significant interactions between these two factors. As in Experiment 5, we averaged across the two measurement conditions in order to reduce modulatory effects caused by the placement of the comparison line.

[4] In all analyses of this study, if Mauchly's (1940) test of sphericity showed a significant deviance (α= 0.25) from equicorrelation for a repeated factor or for a combination of factors including at least one repeated factor, Greenhouse and Geisser's (1959) Epsilon was used to adjust the degrees of freedom for the averaged tests of significance.

were similar for upright and inverted faces[5], which contrasts with the often reported strong inversion effect for processing configuration in face recognition tasks.

3.5 General Discussion

Many previous studies have stressed the importance and orientation-sensitivity of configural processing for recognizing faces. In the present study we investigated the *perception* of configural information in faces and found new and surprising results. Whereas people are very sensitive in detecting configural differences (Bruce et al., 1991; Haig, 1984; Hosie et al., 1988; Kemp et al., 1990) our study shows that configural information is not perceived veridical but is instead overestimated by 11-41 percent. Inversion strongly impairs configural processing in detection and recognition tasks (e.g. Leder & Bruce, 2000; Murray et al., 2000; Rhodes et al., 1993; Schwaninger & Mast, 1999; Searcy & Bartlett, 1996; Sergent, 1984; Young et al., 1987). In contrast, our study revealed that the perception of configural information is much less orientation-sensitive. Moreover, a comparison between overestimations of distances in upright and in 90° rotated faces showed that the HVI affects the perception of the eye-mouth and the inter-eye distance less than it is the case for lines of the same length and thus fails to provide a simple explanation of the large overestimations.
In short, this study revealed a new and large perceptual illusion in faces and indicates that configural processing does not obey the same rules in perceptual tasks as opposed to detection and recognition tasks.

3.6 References

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38,* 109-144.

Coren, S., & Girgus, J.S. (1978). *Seeing is Deceiving: The Psychology of Visual Illusions*. Hillsdale, NJ: Erlbaum.

Fick, A. (1851). *De errore quodam optico asymenetria bulbi effecto.* Marburg: J.A. Kochii.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 32,* 95-112.

Haig, N.D. (1984). The effect of feature displacement on face recognition. *Perception, 13,* 505-512.

Higashiyama, A. (1996). Horizontal and vertical distance perception: The discorded-orientation theory. *Perception & Psychophysics, 58,* 259-270.

Hosie, J.A., Ellis, H.D., & Haig, N.D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception 17,* 461-474.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19,* 531-543.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology, 53,* 513-536.

Mauchly, J.W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics, 11,* 204-210.

[5] However, the small mean difference of 2.8 percent between adjustments of the eye-mouth distance for upright vs. inverted faces was significant when a paired-samples t-test was used (without Bonferroni adjustment for multiple comparisons), $t(11) = 2.56$, $p < .05$.

Murray, J.E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11,* 492-496.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition, 47,* 25-57.

Schwaninger, A., & Mast, F. (1999). Why is face recognition so orientation-sensitive? Psychophysical evidence for an integrative model. *Perception, 28* (Suppl.), 116.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of components and spatial relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904-915.

Sergent, J. (1984). An investigation into component and configurational processes underlying face recognition. *British Journal of Psychology, 75,* 221-242.

Yang, T.L., Dixon, M.W., & Proffitt, D.R. (1999). Seeing big things: Overestimation of heights is greater for real objects than for objects in pictures. *Perception, 28,* 445-467.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configural information in face perception. *Perception, 16,* 747-759.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141-145.

4 Role of featural and configural information in familiar and unfamiliar face recognition

4.1 Abstract

Using psychophysics it was investigated to what extent human face recognition relies on local information in parts (featural information) and on their spatial relations (configural information). This is particularly relevant for biologically motivated computer vision since recent approaches have started considering such featural information. Experiment 7 showed that previously learnt faces could be recognized by human subjects when they were scrambled into constituent parts. This result clearly indicates a role of featural information. Then the blur level was determined that made the scrambled part versions impossible to recognize. This blur level was applied to whole faces in order to create configural versions that by definition do not contain featural information. It was shown that configural versions of previously learnt faces could be recognized reliably. In Experiment 8 these results were replicated for familiar face recognition. Both experiments provide evidence in favour of the view that recognition of familiar and unfamiliar faces relies on featural and configural information. Furthermore, the balance between the two does not differ for familiar and unfamiliar faces. The integrative model of familiar and unfamiliar face recognition proposed in section 2.5 (Figure 7) is discussed as well as implications for biologically motivated computer vision algorithms for face recognition.

4.2 Introduction

Different object classes can often be distinguished using relatively distinctive features like color, texture or global shape. In contrast, face recognition entails discriminating different exemplars from a quite homogeneous and complex stimulus category. Several authors have suggested that such expert face processing is holistic, i.e. faces are meant to be encoded and recognized as whole templates without representing parts explicitly [4,5,6]. In computer vision many face recognition algorithms process the whole face without explicitly processing facial parts. Some of these algorithms have been thought of being particularly useful to understand human face recognition and were cited in studies that claimed faces to be the example for exclusive holistic processing (e.g. [7,8] cited in [9], or the computation models cited in [6], p. 496).

In contrast to holistic algorithms like principal components analysis or vector quantization, recent computer vision approaches have started using local part-based or fragment-based information in faces [1,2,3]. Since human observers can readily tell the parts of a face such algorithms bear a certain intuitive appeal. Moreover, potential advantages of such approaches are greater robustness against partial occlusion and less susceptibility to viewpoint changes.

In the present study psychophysics was used to investigate whether human observers only process faces holistically, or whether they encode and store the local information in facial parts (featural information) as well as their spatial relationship (configural information). In contrast to previous studies, a method was employed that did not alter configural or featural information, but eliminated either the one or the other. Previous studies have often attempted to directly alter the facial features or their spatial positions. However, the effects of such manipulations are not always perfectly selective. For example altering featural information by replacing the eyes and mouth with the ones from another face could also change their spatial relations (configural information) as mentioned in [10]. Rakover has pointed out that altering configuration by increasing the inter-eye distance could also induce a part-change, because the bridge of the nose might appears wider [11]. Such problems were avoided in this study by using scrambling and blurring procedures that allowed investigating the role of featural and configural information separately. The current study extends previous research using these manipulations (e.g. [12,13,14]) by ensuring that each procedure does effectively eliminate configural or featural processing.

4.3 Experiment 7: Unfamiliar face recognition

The first experiment investigated whether human observers store featural information *independent* of configural information. In the first condition configural information was eliminated by cutting the faces into their constituent parts and scrambling them. If the local information in parts (featural information) is encoded and stored, it should be possible to recognize faces above chance even if they are scrambled. In condition 2 the role of configural information was investigated. Previously learnt faces had to be recognized when they were shown as grayscale low-pass filtered versions. These image manipulations destroyed featural information while leaving the configural information intact. In a control condition it was confirmed that performance is reduced to chance when faces are low-pass filtered and scrambled, thus showing that the image manipulations eliminate featural and configural information respectively and effectively.

4.3.1 Participants, materials and procedure

Thirty-six participants, ranging in age from 20 to 35 years voluntarily took part in this experiment. All were undergraduate students of psychology at Zurich University and all reported normal or corrected-to-normal vision.
The stimuli were presented on a 17” screen. The viewing distance of 1 m was maintained by a head rest so that the faces covered approximately 6° of the visual angle. Stimuli were created from color photographs of 10 male and 10 female undergraduate students from the University of Zurich who had agreed to be photographed and to have their pictures used in psychology experiments. All faces were processed with Adobe Photoshop, proportionally scaled to the same face width of 300 pixels and placed on a black background. These intact faces were used in the learning phase (Figure 10a).
The scrambled faces were created by cutting the intact faces into 10 parts, using the polygonal lasso tool with a 2 pixel feather. The number of parts was defined by a preliminary free listing experiment, in which 41 participants listed all parts of a face. The following parts were named by more than 80% of the participants and were used in this study: eyes, eyebrows, nose, forehead, cheeks, mouth, and chin. Four different scrambling versions, which appeared randomly, were used. Each version was arranged

so that no part was situated either in its natural position or in its natural relation to its neighbouring part. The parts were distributed as close to each other as possible, in order to keep the image area approximately the same size as the whole faces (Figure 10b).

The control stimuli were created in three steps. First, all colour information was discarded in the intact faces. Second, the faces were blurred using a Gaussian filter with a sigma of 0.035 of image width in frequency space, which was determined in pilot studies. The formula used to construct the filter in frequency space was $\exp(\frac{-f^2}{2\sigma^2})$.

Figure 10 Sample Stimuli. a) intact face, b) scrambled, c) scrambled-blurred, d) blurred face.

In the third step these blurred faces were cut and scrambled as described above. Figure 10c shows an example of the control faces. The blurred stimuli were created by applying the low-pass filter determined in the control condition to greyscale versions of the intact faces (Figure 10d).

Participants were randomly assigned to one of three groups. Each group was tested in one experimental condition, scrambled, scrambled-blurred, or blurred. Ten randomly selected faces served as target faces and the other 10 faces were used as distractors. In the learning phase the target faces were presented for ten seconds each. After each presented face the screen went blank for 1000 ms. Then the same faces were again presented 10 seconds each in the same order. The faces were presented in a pseudo-random order so that across participants no face appeared at the same position more than twice.

In the experimental phase, 20 faces were presented (10 targets and 10 distractors). Six random orders were created using the following constraints: within each random order no more than three target or distractor faces occurred on consecutive trials and between random orders no face appeared more than once on each position. The same random orders were used for all conditions. Each trial started with a 1000 ms blank followed by a face. The participants were required to respond as fast and as accurately as possible whether the presented face was new (distractor) or whether it had been presented in the learning phase (target) by pressing one of two buttons on a response box. The assignment of buttons to responses was counterbalanced across participants.

4.3.2 Results and discussion

Recognition performance was calculated using signal detection theory [15]. Face recognition performance was measured by calculating d' using an old-new recognition task [16]. This measure is calculated by the formula d' = z(H) – z(FA), whereas H denotes the proportion of hits and FA the proportion of false alarms. A hit was scored when the target button was pressed for a previously learned face (target) and a false alarm was scored when the target button was pressed for a new face (distractor). In the formula z denotes the z-transformation, i.e. H and FA are converted into z-scores (standard-deviation units). d' was calculated for each participant and averaged across each group (Figure 11, black bars).

One sample t-tests (one-tailed) were carried out in order to test the group means *M* against chance performance (i.e. d' = 0). Faces were recognized above chance, even when they were cut into their parts, $M = 1.19$, $SD = 0.58$, $t(11) = 7.07$, $p < .001$. This result suggests that local part-based information has been encoded in the learning phase, which provided a useful representation for recognizing the scrambled versions in the testing phase. These findings are contradictory to the view that faces are only processed holistically [4,5,6,9]. The recognition of blurred faces was also above chance, $M = 1.67$, $SD = 0.82$, $t(11) = 7.044$, $p < .001$. The blur filter used did indeed eliminate all featural information since recognition was at chance when faces were blurred and scrambled, $M = -0.22$, $SD = 1.01$, $t(11) = -.75$, $p = .235$.

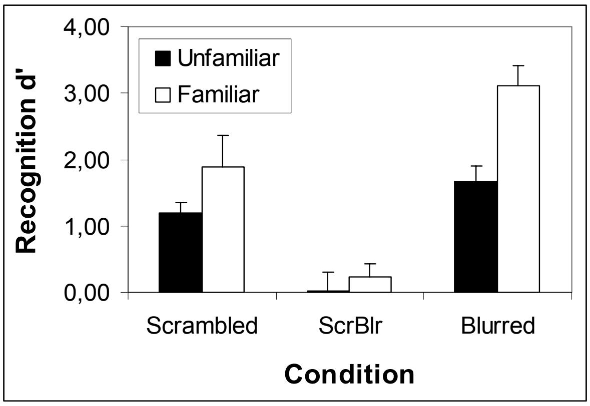

Figure 11 Recognition performance in unfamiliar and familiar face recognition across the three different conditions at test. ScrBlr: scrambled and blurred faces. Error bars indicate standard errors of the mean.

Taken together, these results provide clear evidence for the view that featural and configural information are both important sources of information in face recognition. Furthermore, the two processes do not appear to be arranged hierarchically, as the results show that featural and configural information can be encoded and stored independently of one another[6].

4.4 Experiment 8: Comparison of unfamiliar and familiar face recognition

The results of Experiment 7 challenge the hypothesis that faces are only processed holistically. At the same time these results suggest that for unfamiliar face recognition in humans separate representations exist for featural information and configural information. The aim of Experiment 8 was to investigate whether the same is true for *familiar* face recognition. Moreover, by comparing recognition performance from Experiment 7 and Experiment 8 the question was addressed whether there is a shift in processing strategy from unfamiliar to familiar face recognition. Neuropsychological evidence suggests a dissociation between familiar face recognition and unfamiliar face matching [17,18], and experimental evidence suggests that familiar face recognition relies more heavily on the processing of inner areas of the face than does unfamiliar face recognition [19]. However, previous studies have found no evidence for a change in the balance between featural and configural processing as faces become more familiar [20,12]. This study aimed to clarify this issue using a design that carefully controls the available featural and configural cues in the input image. Furthermore, in contrast to previous studies this study used the same faces in both experiments to eliminate other potential confounds with familiarity.

4.4.1 Participants, materials and procedure

Thirty-six participants ranging in age from 20 to 35 years took part in this experiment for course credits. All were undergraduate students of psychology at Zurich University

[6] It is worth noting, however, that just because featural and configural processing *can* be recognized independently of one another, does not prove that the two don't interact when both are available (e.g. [5])

and were familiar with the target faces. All reported normal or corrected-to-normal vision. The materials and procedure were the same as in Experiment 7. The stimuli were also the same, but all the targets were faces of fellow students and thus familiar to the participants. All distractor faces were unfamiliar to the participants.

4.4.2 Results and discussion

The same analyses were carried out as in Experiment 7. Again, one-sample t-tests (one-tailed) revealed a significant difference from chance (i.e. d' > 0) for recognizing scrambled faces, $M = 2.19$, $t(11) = 4.55$, $p < .001$, and blurred faces, $M = 2.92$, $t(11) = 9.81$, $p < .001$. As in Experiment 7, scrambling blurred greyscale versions provided a control condition for testing whether the blur filter used did indeed eliminate all local part-based information. This was the case – faces could no longer be recognized when they were blurred and scrambled, $M = 0.19$, $t(11) = 0.94$, $p = .184$.

In short, the results of Experiment 8 replicated the clear effects from Experiment 7 and suggest an important role of local part-based and configural information in both unfamiliar and familiar face recognition. By comparing recognition performance from both experiments (Figure 11) the question was addressed to what extent familiar and unfamiliar face recognition differ quantitatively (e.g. generally a better performance when faces are familiar) or qualitatively (e.g. better performance for familiar faces using more accurate configural processing). To this end, a two-way analysis of variance (ANOVA) was carried out with the data from the scrambled and blurred conditions of Experiments 7 and 8 with familiarity (familiar vs. unfamiliar) and condition (scrambled vs. blurred) as between-subjects factors. There was a main effect of familiarity, $F(1,42) = 12.80$, $MSE = 13.48$, $p < .01$, suggesting that familiar faces are more reliably recognized than unfamiliar faces (quantitative difference). There was also a main effect of condition, $F(1,42) = 6.7$, $MSE = 7.05$, $p < .05$, indicating that blurred faces were better recognized than scrambled faces. The relative impact of blurring and scrambling did not differ between the two experiments, since there was no interaction between condition and familiarity, $F(1,42) = 1.02$, $MSE = 1.08$, $p = 0.32$. This results suggests that there are no qualitative differences between familiar and unfamiliar face recognition on the basis of configural and featural information. In both cases both types of information are of similar importance.

4.5 General discussion

In the present paper the role of local part-based information and their spatial interrelationship (configural information) was investigated using psychophysics. It was found that human observers process familiar and unfamiliar faces by encoding and storing configural information as well as the local information contained in facial parts. These results challenge the assumption that faces are processed only holistically and suggest a greater biological plausibility for recent machine vision approaches in which local features and parts play a pivotal role (e.g. [1,2,3]).

Neurophysiological evidence supports part-based as well as configural and holisitic processing assumptions. In general, it has been found that cells responsive to facial identity are found in inferior temporal cortex while selectivity to facial expressions, viewing angle and gaze direction can be found in the superior temporal sulcus [21, 22]. For some neurons, selectivity for particular features of the head and face, e.g. the eyes

and mouth, has been revealed [22,23,24]. Other groups of cells need the simultaneous presentation of multiple parts of a face and are therefore consistent with a more holistic type of processing [25,26]. Finally, Yamane et al. [27] have discovered neurons that detect combinations of distances between facial parts, such as the eyes, mouth, eyebrows, and hair, which suggest sensitivity for the spatial relations between facial parts (configural information).
The model ideas presented in section 2.5 are very fruitful for integrating the above mentioned findings from psychophysics, neurophysiology and computer vision (Figure 12). Faces are first represented by a metric representation in primary visual areas corresponding to the perception of the pictorial aspects of a face. Further processing entails extracting local part-based information and spatial relations between them in order to activate featural and configural representations in higher visual areas of the ventral stream, i.e. face selective areas in temporal cortex[7]. In a recent study, repetition priming was used in order to investigate whether the outputs of featural and configural representations converge to the same face identification units [28, see also chapter 5]. Since priming was found from scrambled to blurred faces and vice versa it is proposed that the outputs of featural and configural representations converge to the same face identification units.

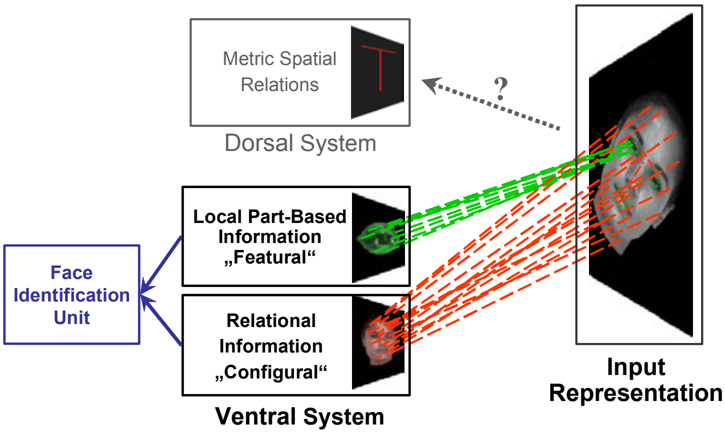

Figure 12 Integrative model for unfamiliar and familiar face recognition.

4.6 References

1. Heisele, B., Serre, T., Pontil, M., Vetter, T., and Poggio, T. (2001). Categorization by learning and combining object parts. NIPS proceedings.
2. Lee, D.D., & Seung, S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788-791.
3. Ullman, S., & Sali, E. (2000). Object Classification Using a Fragment-Based Representation. BMCV 2000. Lecture Notes in Computer Science, 1811, pp. 73-87. Berlin: Springer.
4. Tanaka J. W. & Farah, M. J. (1991). Second-order relational properties and the inversion-effect: Testing a theory of face perception. Perception & Psychophysics, 50, 367-372.
5. Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology, 79, 471-491.
6. Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception and Performance, 21, (3), 628-634.

[7] Although a role of the dorsal system in encoding of metric spatial relations has been proposed for object recognition it remains to be investigated, whether it does play a role for the processing of configural information in faces.

7. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., Malsburg, v.d., C., Würtz, R.P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42, 300-311.
8. Wiskott, L., Fellous, J.M., Krüger, N., & von der Malsburg, C. (1997). Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 775-779.
9. Biedermann, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society of London, B, 352, 1203-1219.
10. Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? Cognition, 47, 25-57.
11. Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. British Journal of Psychology, 93, 1-30.
12. Collishaw, S.M., Hole G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29, 893-910.
13. Davidoff, J., & Donnelly, N. (1990). Object superiority: a comparison of complete and part probes, Acta Psychologica, 73 1990 225–243.
14. Sergent J. (1985). Influence of task and input factors on hemispheric involvement in face processing. Journal of Experimental Psychology: Human Perception and Performance, 11(6), 846-61.
15. Green, D.M., & Swets, J.A. (1966). Signal detection theory and psychophysics. New York: Wiley.
16. McMillan, N.A., & Creelman, C.D. (1992). Detection theory: A user's guide. New York: Cambridge University Press.
17. Benton AL, 1980. The neuropsychology of facial recognition. American Psychologist, 35, 176-186.
18. Malone D.R., Morris H.H., Kay M.C., Levin, H.S., 1982. Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. Journal of Neurology, Neurosurgery, and Psychiatry, 45, 820-822.
19. Ellis, H.D., Shepherd, J.W., Davies G.M., 1979. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception, 8, 431-439
20. Yarmey, A.D., 1971. Recognition memory for familiar "public" faces: Effects of orientation and delay. Psychonomic Science, 24, 286-288.
21. Hasselmo, M.E., Rolls, E.T., & Baylis, C.G. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Experimental Brain Research, 32, 203-218.
22. Perret, D.I., Hietanen, J.K., Oram, M.W., & Benson, P.J. (1992). Organization and functions of cells in the macaque temporal cortex. Philosophical Transactions of the Royal Society of London, B, 335, 23-50.
23. Perret, D.I., Rolls, E.T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Experimental Brain Research, 47, 329-342.
24. Perret, D.I., Mistlin, A.J., & Chitty, A.J. (1987). Visual neurones responsive to faces. Trends in Neuroscience, 10, 358-364.
25. Perret, D.I., & Oram, M.W. (1993). Image Vis. Comput., 11, 317-333.
26. Wachsmuth, E., Oram, M.W., & Perret, D.I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cerebral Cortex, 4, 509-522.
27. Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? Experimental Brain Research, 73, 209-214.

28. Schwaninger, Lobmaier, & Collishaw (2002). Role and interaction of featural and configural processing in face recognition. Vision Sciences Society, 2nd annual meeting, Sarasota, Florida, May 10-15, 2002.

5 Convergence of configural and featural processing in face recognition – evidence from repetition priming

5.1 Abstract

In previous chapters converging evidence was presented suggesting that face recognition involves both configural and featural processing. In this study, repetition priming was used to investigate the relationship between these processes: do scrambled faces (containing primarily featural information) prime blurred faces (containing primarily configural information), and vice versa? Three groups of participants saw a set of familiar faces twice. Group NN saw the faces in unmanipulated versions both times; group SB saw them scrambled at time 1 and blurred at time 2; group BS saw them blurred at time 1 and scrambled at time 2. Significant repetition priming effects were found in all three conditions. The amount of priming was as large for group SB as it was for group NN. These results are contrary to a dual-process model of face recognition in which featural and configural analyses provide independent inputs to face recognition. Instead, this result is consistent with configural and featural analyses converging to common face representations used for identification.

5.2 Introduction

It is well known that faces are not only perceived and recognised in terms of their individual features. The spatial configuration of the elements and the shape of the face also provide important cues to the identity of a face. The distinction between configural and featural processing is influential in many accounts of face recognition (e.g.: Sergent, 1984; Diamond & Carey, 1986; Rhodes et al., 1993; Searcy & Bartlett, 1996; Leder & Bruce, 1998), and numerous studies provide evidence that configural and featural information are both important in face recognition. People are very sensitive to changes in the spatial relationships of the facial features (Haig, 1984). Recognition survives when faces are blurred to such an extent that local featural information can no longer be perceived (e.g. Harmon, 1973; Costen et al., 1994). Finally, component parts of composite faces are hard to isolate from their whole face context (Young et al., 1987). However, local featural information also plays an important role – the recognition of faces on the basis of jumbled-up or isolated facial features is found to be above chance levels (e.g. Tanaka & Farah, 1993; Collishaw & Hole, 2000).
Evidence that there is a qualitative distinction between configural and featural processing comes from various sources. First, each process appears differentially affected by inversion. In particular, featural cues are equally well perceived in upside down as upright faces, whilst inversion leads to a significant impairment in the perception and recognition of configural information (e.g.: Endo, 1986; Searcy & Bartlett, 1996; Murray et al. 2000; see also Yin, 1969; Valentine, 1988). Studies have shown contrasting patterns of lateralisation for configural and featural processing. In particular, there appears to be a right-hemisphere advantage for the processing of upright faces (Rhodes, 1985), for the processing of low spatial frequency information

(Sergent, 1986) and for whole-face processing (Rossion et al., 2000). Studies of developmental disorders provide evidence for a selective impairment of configural processing in autism and Williams syndrome (see Elgar & Campbell, 2001). Together, this evidence supports the view that featural and configural cues are based on different representations. However, relatively little is known about the nature of the relationship between configural and featural processing. It is unclear whether they are hierarchically arranged, independent of one another, or whether the outputs of configural and featural processes converge.

An important tool for the understanding of the nature of these two processes has been the development of behavioural marker tasks that can isolate the contribution of each. Attempts at directly manipulating either featural or configural information have proved difficult to interpret, because the effects of such alterations are often ambiguous. For example, changing the shape and size of individual features will also affect the relationship between these features and the rest of the face (Rhodes et al., 1993). Similarly, changing the spatial relations in the face may lead to changes in the shapes of individual features (e.g.: changing the vertical position of the mouth changes the size of the chin). More general manipulations, however, have proved fruitful in controlling for the effects of one or the other process.

The experiments presented in chapter 4 (Experiments 7 and 8) provide a solution to this problem because the method applied selectively altered featural and configural information. By scrambling faces their spatial interrelationship is changed, which eliminates the availability of configural information. Blurring faces using a low pass filter destroys local details of the parts and therefore eliminates featural information. The study provided strong support for a distinction between featural and configural processing, demonstrated that each form of processing *can* occur independently of the other, and showed that the manipulations that were used effectively distinguish between configural and featural forms of information.

The aim of this study is to further examine the nature of the relationship between configural and featural processing. Several considerations are important. First, whilst previous studies show that featural and configural cues *can* be processed independently of one another, this does not mean that in normal face recognition, features and configuration are processed independently. Recognition is optimal when both forms of information are available, and several studies suggest that the whole face context enables more efficient processing of individual face features (e.g. Tanaka & Farah, 1993). Secondly, the results of chapter 4 are contrary to a hierarchical model – they suggest that featural processing does not rely on configural information, whilst configural processing does not depend on fine-detail featural information. Two hypotheses about the relationship between featural and configural processing can be formulated. According to the first, featural and configural processing proceed independently of one another at all stages of face perception and representation. A second hypothesis suggests that processing converges, and that information about features and their configuration are combined into unitary holistic representations. The aim of the experiments presented here is to use a repetition priming paradigm to help understand the relationship between configural and featural processing.

Repetition priming is the phenomenon where the repeated presentation of a stimulus leads to an observer recognising it more efficiently. Repetition priming has been found for a variety of stimuli, including words (Morton, 1979), objects (Warren & Morton,

1982) and faces (Bruce & Valentine, 1985). Repetition priming effects, along with the related phenomenon of associative priming[8], have proved important for understanding the cognitive architecture of the face recognition system (Bruce and Young, 1986). According to the Bruce and Young model, face recognition proceeds independently of other forms of face perception (e.g. expression and facial speech analysis). Viewpoint-independent codes are derived from initial perceptual encoding and matched against stored representations held at 'face recognition units (FRU). Semantic information about people is stored at 'Person Identity Nodes' (PIN) which can be activated by outputs of FRUs or via other routes (e.g. voices and names). Two possible mechanisms have been proposed to explain repetition priming effects in face recognition. According to Bruce and Young (1986), perceptual representations of faces are held at 'Face Recognition Units' (FRU) that have a certain resting level of activation, while increased activation of the unit signals familiarity. Over time, activation at the unit decays, and the unit is returned to its resting level. Repetition priming effects occur when presentation of a face is repeated before the associated FRU has returned to its resting level. An alternative proposal is that repetition priming occurs because the recognition of a familiar face activates and strengthens the pathway between the face recognition unit and an associated Person Identity Node (Burton et al, 1990; Johnston and Barry, 2001). This proposal may be better able to account for the long-lasting nature of repetition priming effects that have now been demonstrated – in some cases over a matter of months (e.g. Maylor, 1998). By demonstrating that long-lasting repetition priming effects are modality specific (voices, names, and faces do not prime each other[9], researchers have provided strong evidence for independent access of person-specific identity through these three routes (e.g. Bruce & Valentine, 1985; Ellis et al., 1987; Ellis et al., 1997a). In contrast, different views of the same face do elicit strong repetition priming effects (Bruce and Valentine, 1985; Ellis et al., 1987; Johnston & Barry, 2001), suggesting that FRUs accumulate input from different views (Bruce and Young, 1986).

The current experiment uses the repetition priming paradigm to help understand the relationship between configural and featural processing. The experiment will examine the extent to which the recognition of blurred faces at time 1 primes the recognition of scrambled faces at time 2 (and vice versa). The experimental hypothesis is that priming effects will be found, providing evidence in support of a convergent model of face perception and recognition. The alternative hypothesis, that there is no priming between blurred and scrambled face recognition, would be consistent with the view that there are separate representations of featural and configural information at all stages of perceptual processing. According to this view, each would provide a separate route to our store of information about people at the level of personal identity - an effect analogous to the lack of priming between face, name and voice recognition. In order to

[8] Associative priming has been demonstrated by Bruce & Valentine (1986). They showed that face recognition is faster if preceded by a related face (e.g. Princess Diana then Prince Charles). Associative priming has been explained in terms of increased excitation of units that store semantic information shared by related faces. Associative priming is short-lived and in one study was shown to be abolished by introducing even one intervening item between prime and target (Ellis, 1992).

[9] Short-lived cross-modal priming effects may arise, possibly due to the effects of associative priming (e.g. Calder & Young, 1996; Ellis et al., 1997a).

rule out short-lived associative priming effects, primes and targets will be presented in separate blocks, separated by numerous intervening items.

Previous studies provide a number of important insights, although none have been designed to address this particular question. Various studies have examined the extent to which whole face recognition is primed by prior exposure to part or whole faces. These studies have shown that both types of prime induce gains in performance when compared to unprimed recognition, (Brunas et al, 1990; Brunas-Wagstaff et al., 1992; Johnston et al., 1996; Ellis et al., 1997b). Whilst whole face priming in some studies produces effects that exceed those of part-face priming (e.g. Ellis et al., 1997b), other studies have found no difference between the two conditions (e.g. Brunas-Wagstaff et al., 1992; Johnston et al., 1996). Instead, these studies find that the extent of the repetition priming effect depends more on whether the prime face is recognised than on its similarity to the test face. One problem for many of these studies concerns the fact that priming effects may in part reflect the effects of overlapping pictorial information (and thus may reflect the operation of non-face-specific visual processing). This concern has been addressed by Ellis et al., (1997b) who demonstrated the existence of non-pictorial priming from one part of the face to a different region of the face.

The priming effects in the studies discussed so far do not address the relationship between configural and featural processing in face recognition. Most studies have examined the priming of whole face recognition, whilst even the part-face conditions have not unambiguously distinguished between the effects of configuration and feature. The part-to-part priming in the Ellis et al (1997b) study is consistent with both independent and convergent models of face recognition. Indeed, Ellis et al (1997b) point out that their results indicate that repetition priming effects can occur at several points within the face recognition system. They argue that some priming occurs due to the strengthening of links between sets of pre-FRU feature detectors[10], whilst some priming reflects the strengthening between FRUs and PINs. One further study of relevance is that by Arguin & Saumier (1999, cited in Saumier et al., 2001). They found that the size of part to whole face priming increases exponentially with the number of parts shown, suggesting that emerging configural information plays an important role. Further support for this view comes from the result that single part priming was only found when the part was presented in a (generic) whole face context.

None of the studies reviewed has directly addressed the question of whether priming effects are found between configural and featural facial information. Our study builds on the findings from chapter 4, by assuming that scrambled and blurred faces activate separate representations based on parts (featural or component information) and the holistic aspects of a face (configural information).If scrambled faces presented at time 1 are able to prime the recognition of blurred faces presented at time 2 or vice versa, then this demonstrates that configural and featural processing are not independent of one another. Such results would instead support a convergent model of face recognition.

[10] Ellis et al (1997b) suggest that feature detectors may take the form of different global descriptions of the face.

5.3 Experiment 9: Repetition priming study

5.3.1 Participants and design

Forty-eight people were recruited. 15 participants were male and 33 were female, and their mean age was 25.1 years. All were naive about the purpose of the study. Individuals were randomly assigned to one of three groups. As Figure 13 shows, priming effects in each condition were measured using the same two-phase design. First, participants were asked to discriminate well-known celebrities from unknown distractor faces. These celebrities functioned as the primes for the second phase of the study. In phase 2, the speed of recognition for new faces and for previously shown faces was compared. In group 1 (*N-N*, the control group), priming of normal faces was assessed. Unmanipulated faces were shown in phase 1 and in phase 2. In group 2 (*S-B*), scrambled faces were used as primes in phase 1, and blurred versions were shown in the second phase. For group 3 (*B-S*), this order was reversed, and blurred faces were used to prime scrambled faces.

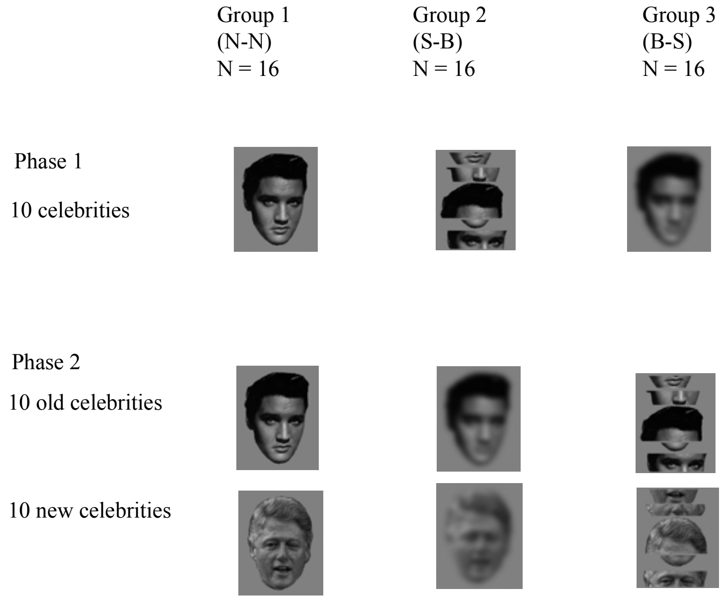

Figure 13 Sample stimuli and illustration of design.

5.3.2 Materials and apparatus

The study used twenty famous targets and thirty unfamiliar distractor faces. All the faces were of clean-shaven, Caucasian, adult men. All targets were well known in the UK, and included TV personalities, sportsmen, politicians and royalty. Distractor faces were broadly matched with the target faces on the basis of the following criteria: view, expression, hair colour, hair length, and age. Furthermore, these faces were famous in the Netherlands, though not in the UK. The rationale for using Dutch celebrities as distractor faces was to gain additional control over extraneous factors that might indicate that a face was a celebrity even when they are unfamiliar to a participant (e.g. image source, expression, pose, make up, etc).

Each face was scanned from a magazine and subsequently prepared in Adobe Photoshop by replacing all the background information with a uniform grey background, orienting the face upright (if necessary), and scaling it to a standard size across the width of the image (450 pixels). Blurred versions of the faces were prepared using a Gaussian blur with a filter radius of 10 pixels, as implemented in Photoshop. Scrambled faces were prepared by dividing them into five horizontal strips, and re-arranging these in the following order from top to bottom: a) mouth; b) nose; c) forehead and hair; d) chin; e) eyes. Examples of target stimuli used in this study are shown in Figure 13. Stimuli were displayed using SuperlabPro 1.05. Participants made their choice (famous or not famous) using the left and right mouse buttons, and the time between the onset of the stimulus and the mouse button press by the participant were recorded. Images measured 90mm by (approximately) 120mm. All participants were tested at a viewing distance of 50-60cm. The images thus subtended about 8.06° horizontally.

5.3.3 Procedure

Participants in all groups were tested using twenty faces (ten target and ten distractor faces) in phase 1 of the experiment. Faces were shown one at a time, and participants were instructed to respond as quickly and as accurately as possible by pressing the left or the right mouse button to indicate whether they thought that a particular face was famous or not. Following each response, there was a delay of one second before the onset of the next trial. The order of presentation of the stimuli was randomised for each participant. In phase 2 of the test, participants were shown 20 targets - ten previously shown and ten new faces, together with 20 distractor faces. Sets of primed and unprimed target faces were counter-balanced across participants. Group *N-N* received normal unmanipulated faces in both phases of the experiment, group *S-B* were tested using scrambled faces in phase 1 and blurred faces in phase 2, whilst group *B-S* received blurred faces in phase 1 and scrambled faces in phase 2. Before the start of the second experimental phase, participants were reminded of the importance of responding quickly and accurately.

5.3.4 Statistical analysis

The main focus was on possible differences between reaction times for primed and unprimed faces in the second phase of the study, and whether similar effects were apparent for each group. Analyses of reaction time (RT) were restricted to latencies for correct responses to target faces, as these are the responses for which possible priming effects may be expected. To reduce the error that may result from outliers in a RT paradigm (Ulrich and Miller, 1994), M+2SD cut-offs were calculated for each group, separately for primed and unprimed faces. In total, 5.2% of responses were excluded. Two-tailed dependent t-tests were used to test for differences between unprimed and primed responses for each group. Mixed-design analyses of variance tested for group differences in the size of priming effects.

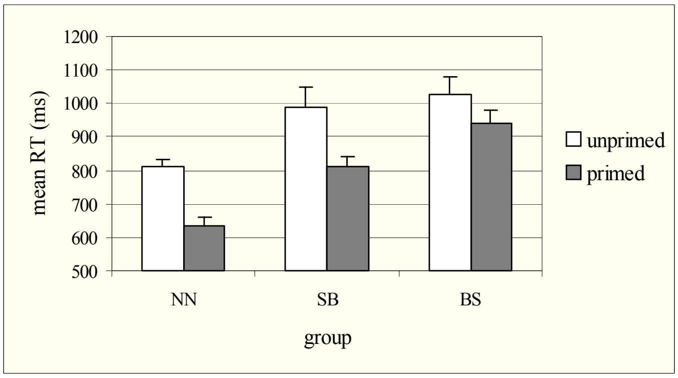

Figure 14 Repetition priming of celebrity face recognition. Mean reaction times (+1SE) in test phase for primed and unprimed faces separately for each group (NN - normal faces in prime and test phase, SB - scrambled faces in prime phase and blurred faces in test phase, BS - blurred faces in prime phase and scrambled faces in test phase).

5.3.5 Results

Figure 14 shows mean RTs (and standard errors) for correct identifications of primed and unprimed target faces separately for each group. There were strong priming effects in all three conditions. Participants in group *N-N* were 176ms faster in recognising normal primed celebrities, $t(15) = 8.20$, $p < 0.001$. Participants in group *S-B* were 178ms faster in recognising blurred faces that were primed by scrambled exemplars $t(15) = 3.54$, $p = 0.003$. Group *B-S* showed an 88ms advantage for the recognition of scrambled faces primed by the previous recognition of blurred versions of the faces, $t(15) = 2.34$, $p = 0.03$. A mixed design analysis of variance found significant main effects of priming, $F(1, 45) = 44.49$, $p < 0.001$, and group, $F(2, 45) = 13.94$, $p < 0.001$, but no interaction between the two, $F(2, 45) = 1.83$, $p = 0.17$.

5.3.6 Discussion

The results of this experiment showed repetition priming effects in all three conditions. Participants were 176ms faster at recognising primed intact faces, 178ms faster at recognising blurred faces previously shown in scrambled form, and 88ms faster when recognising scrambled faces primed by blurred faces[11]. These results imply that common cognitive mechanisms underlie the processing of scrambled and blurred faces at some stage, despite the fact that the prime and test faces had little pictorial information in common.

The main aim of the study was to assess possible models of the relationship between featural and configural processing. Repetition priming effects imply the activation of common representations, and the results therefore provide evidence against an independent model of configural and featural processing. Whilst the possibility that the images contain some overlapping information must be considered (e.g. some residual configural information is retained in scrambled images to prime later configural processing of the blurred faces), there are several reasons for thinking that it cannot explain the repetition priming effects demonstrated here. The magnitude of the priming effects was large, and scrambled to blurred priming was as great as priming for the repeated presentation of identical pictures of normal intact faces.

An additional issue relates to the extent to which associative priming influenced the results obtained in this study. Associative priming occurs when retrieval of semantic and name information associated with a face presented at time 1 leads to more efficient retrieval of this information at time 2. Confounding effects arising as a result of associative priming effects are thought unlikely for several reasons. Previous research on the nature of associative priming suggests that it is a very short-lived phenomenon (e.g. Bruce and Valentine, 1986; Ellis, 1992; Calder and Young, 1996). For example, Bruce and Valentine (1986) found that in a series of faces the presentation of a famous face primed the recognition of the next face if it was related (e.g. Princess Diana followed by Prince Charles). However, no priming occurred when there was an unrelated intervening item. Similarly, priming between voices, names and faces also only occurs at very short intervals (e.g. Young et al., 1994; Ellis et al., 1997a). In line with previous studies with a focus on repetition priming, the current experiment has used separate blocks of prime and target faces to control for possible associative priming effects.

---

[11] It is unclear why priming from blurred to scrambled face recognition should be smaller than for the other two conditions. However, there was no significant interaction between priming and condition. The key point for the present purpose is that priming effects were significant in all three conditions.

The current study confirms the prevailing view from previous research that different encounters with a face will generally lead to activation of some common representation at some stage in processing. Previous studies have demonstrated priming between different views of the face (e.g. Bruce and Valentine, 1985; Johnston and Barry, 2001), between part and whole faces (e.g. Brunas-Wagstaff et al., 1992), and between complementary sets of facial features (Ellis et al 1997b). The findings in this study demonstrate that strong repetition priming also occurs when featural information is shown in the priming phase and configural information in the test phase, or vice versa. The results are consistent with a convergent model of face recognition in which faces are initially processed using separable configural and featural mechanisms, and where outputs converge and accumulate to form common representations, i.e. face identification units. According to this view, activation of these units, or of pathways leading from these units, results in the priming between scrambled and blurred faces found in this study.

5.4 Conclusions

This study has demonstrated that scrambled and blurred faces are able to prime one another, suggesting that local featural information and configural information activate common representational mechanisms. This finding constrains the range of possible models of the relationship between configural and featural processing by ruling out a dual-route independent access model of face recognition. The results are consistent with a convergent model of face recognition where configural and featural analyses activate common unitary representations of faces.

5.5 References

Bruce, V., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76*, 363-383.

Bruce, V., & Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology, 38A*, 125-150.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305-327.

Brunas, J., & Young, A.W., & Ellis, A.W. (1990). Repetition priming from incomplete faces: Evidence for part-to-whole completion. *British Journal of Psychology, 81*, 43-56.

Brunas-Wagstaff, J., Young, A.W., & Ellis, A.W. (1992). Repetition priming follows spontaneous but not prompted recognition of familiar faces. *Quarterly Journal of Experimental Psychology, 44*, 423-454.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*, 361-380.

Calder, A.J., & Young, A.W. (1996). Self-priming: A short term benefit of repetition. *Quarterly Journal of Experimental Psychology*, *49A,* 845-861.

Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29*, 893-910.

Costen, N.P., Parker, D.M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*, 129-146.

Costen, N.P., Parker, D.M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics, 58*, 602-612.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Elgar, K., & Campbell, R. (2001). Annotation: The cognitive neuroscience of face recognition. Implications for developmental disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 705-717.

Ellis, A.W. (1992). Cognitive mechanisms of face processing. *Philosophical Transactions of the Royal Society, B335*, 113-119.

Ellis, A.W., Burton, A.M., Young, A.W., & Flude, B.M. (1997b). Repetition priming between parts and wholes: Tests of a computational model of familiar face recognition. *British Journal of Psychology, 88*, 579-608.

Ellis, A.W., Young, A.W., Flude, B.F., & Hay, D.C. (1987). Repetition priming of face recognition. *Quarterly Journal of Experimental Psychology, 39A,* 193-210.

Ellis, H.D., Jones, D.M., & Mosdell, N. (1997a). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88*, 143-156.

Endo, M. (1986). Perception of upside-down faces: an analysis from the viewpoint of cue saliency. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A. Young (Eds.) *Aspects of Face Processing.* Dordrecht: Martinus Nijhoff.

Haig, N.D. (1984). The effect of feature displacement on face recognition. *Perception, 13*, 505-512.

Harmon, L.D. (1973). The recognition of faces. *Scientific American, 229*, 70-83.

Johnston, R.A., & Barry, C. (2001). Best face forward: Similarity effects in repetition priming of face recognition. *Quarterly Journal of Experimental Psychology, 54A***,** 383-396.

Johnston, R.A., Barry, C., & Williams, C. (1996). Incomplete faces don't show the whole picture: Repetition priming from jumbled faces. *Quarterly Journal of Experimental Psychology, 49A*, 596-615.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology A, 51*, 449-473.

Maylor, E.A. (1998). Retrieving names in old age: Short- and (very) long-term effects of repetition. *Memory and Cognition*, *26*, 309-319.

Morton, J. (1979). Facilitation in word recognition: Experiments causing changes in the logogen model. In P.A. Kolers, M. Wrolstad, & H. Bouma (Eds.) *Processing of visible language*. New York: Plenum.

Murray, J.E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11(6)*, 492-496.

Rhodes, G. (1985). Lateralized processes in face recognition. *British Journal of Psychology, 76*, 249-271.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition, 47*, 25-57.

Rossion, B., Dricot, L., Devolder, A., Bodart, J.M., Crommelinck, M., de Gelder, B., & Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. Journal of Cognitive Neuroscience, 12*, 793-802*

Saumier, D., Arguin, M., & Lassonde, M. (2001). Prosopagnosia: A case study involving problems in processing configural information. *Brain and Cognition, 46*, 255-316.

Schwaninger, A., Collishaw, S.M., & Lobmaier, J. (in press). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science,* Heidelberg: Springer.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904-915.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*, 221-242.

Sergent, J. (1986). Microgenesis of face perception. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A. Young (Eds.) *Aspects of Face Processing.* Dordrecht: Martinus Nijhoff.

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*, 225-245.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34-80.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology, 73*, 117-130.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Young, A.W., Flude, B., Hellawell, D.J., & Ellis, A.W. (1994). The nature of semantic priming effects in the recognition of familiar faces. *British Journal of Psychology, 85*, 393-412.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception*, *16*, 747-759.

6 View-based recognition of faces in man and machine: re-visiting inter-extra-ortho

6.1 Abstract

For humans, faces are highly overlearned stimuli, which are encountered in everyday life in all kinds of poses and views. Using psychophysics we investigated the effects of viewpoint on human face recognition. The experimental paradigm is modeled after the inter-extra-ortho experiment using unfamiliar objects by Bülthoff and Edelman [5]. The results in this study show a strong viewpoint effect for face recognition, which replicates the earlier findings and provides important insights into the biological plausibility of view-based recognition approaches (alignment of a 3D model, linear combination of 2D views and view-interpolation). Human recognition performance is compared to a novel computational view-based approach [29] and improvements of view-based algorithms using local part-based information are discussed.

6.2 Introduction

According to Marr [16] human object recognition can be best understood by algorithms that hierarchically decompose objects into their parts and relations in order to access an object-centered 3D model. Based on the concept of nonaccidental properties [14], Biederman proposed in his recognition by components (RBC) theory [1], that the human visual system derives a line-drawing-like representation from the visual input, which is parsed into basic primitives (geons) that are orientation-invariant. Object recognition would be achieved by matching the geons and their spatial relations to a geon structural description in memory. This theory has been implemented in a connectionist network that is capable of reliably recognizing line drawings of objects made of two geons [11].
In object recognition, view-based models have often been cited as the opposite theoretical position to the approaches by Marr and Biederman[12]. Motivated by the still unsolved (and perhaps not solvable) problem of extracting a perfect line drawing from natural images different view-based approaches have been proposed. In this paper three main approaches are considered: Recognition by alignment to a 3D representation [15], recognition by the linear combination of 2D views [25], and recognition by view interpolation (e.g., using RBF networks [19]). What these approaches have in common is that they match viewpoint *dependent* information as opposed to viewpoint *invariant* geons.
The biological plausibility of these models has been investigated by comparing them to human performance for recognizing paper clip and amoeboid like objects [5,7]. In contrast to those stimuli, faces are highly overlearned and seen in a vast variety of different views and poses. Therefore, it was investigated whether a) human face

[12] However, it is interesting that Biederman and Kalocsai [3] point out that face recognition – as opposed to object recognition – cannot be understood by RBC theory mainly because recognizing faces entails processing holistic surface based information.

recognition shows similar effects of viewpoint and b) by which of these view-based approaches face recognition can be best understood. Human recognition performance was then compared to another view-based framework, namely the feature matching approach based on the framework introduced in [29]. Based on the results the role of parts and their interrelationship are discussed from a view-based perspective and contrasted to the models proposed in [1,11].

6.3 Experiment 10: View-based recognition of faces by humans

6.3.1 Participants, method and procedure

Ten right-handed undergraduates (five females, five males) from the University of Zürich volunteered in this study. The face stimuli were presented on a 17" CRT screen. The viewing distance of 1 m was maintained by a head rest so that the faces covered approximately 6° of the visual angle. Twenty male faces from the MPI face database [4] served as stimuli.

The experiment consisted of a learning and a testing phase. Ten faces were randomly selected as distractors and the other 10 faces were selected as targets. During learning, the target faces were shown oscillating horizontally ±5° around the 0° and the 60° extra view (see Figure 15). The views of the motion sequence were separated by 1° and were shown 67 ms per frame. The oscillations around 0° started and ended always with the +5° view, the oscillations around 60° started and ended always with the +55° view. Both motion sequences lasted 6 sec, i.e. 4 full back-and forth cycles. For half the faces the 0° sequence was shown first, for the other half of the faces the 60° sequence was shown first. The order of the ten faces was counterbalanced across the ten participants. After a short break of 15 min the learning block was repeated and for each face the order of the two motion sequences was reversed.

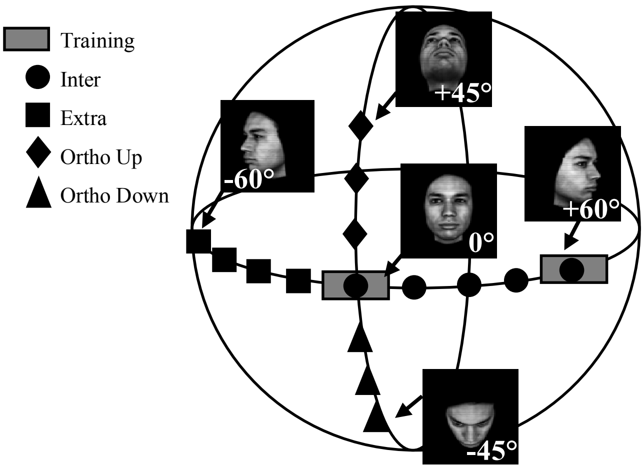

Figure 15 Training occurred at 0° ±5° (frontal view) and 60° ±5° (side view). Testing was performed for 15 views separated by 15°. The four testing conditions are labeled (inter, extra, ortho up, ortho down).

In the testing phase, the subjects were presented with static views of the 10 target and the 10 distractor faces. The faces were shown in blocks of 20 trials in which each face was presented once in a random order. The test phase contained 300 trials and each face was presented once in each of the 15 angles depicted in Figure 15. Each trial started with a 1000 ms fixation cross followed by the presentation of a face. Participants were instructed to respond as fast and accurately as possible whether the presented face had been shown in the learning face (i.e. it was a target) or whether it was a distractor by pressing the left or right mouse button. On each trial, the faces were presented until the button press occurred. The assignment of buttons to responses was counterbalanced across participants.

6.3.2 Results and discussion

Signal detection theory was used to measure recognition performance. The relevant measure is d' = z(H) – z(FA), whereas H equals the hit rate, i.e. the proportion of correctly identified targets, and FA the false alarm rate, i.e. the proportion of incorrectly reporting that a face had been learned in the learning phase. H and FA are converted into z-scores, i.e. to standard deviation units. Individually calculated d' values were subjected to a two-factor analysis of variance (ANOVA) with condition (extra, inter, orthoUp, orthoDown) and amount of rotation (0, 15, 30, 45) as within subjects factors. Mean values are shown in Figure 16.

Recognition d' was dependent on the condition as indicated by the main effect of this factor, $F(3, 27) = 23.1$, $MSE = .354$, $p < .001$. There was also a main effect of amount of rotation, $F(3, 27) = 10.93$, $MSE = 1.500$, $p < .001$. The effect of rotation was different across conditions as indicated by the interaction between amount of rotation and condition, $F(9,81) = 3.30$, $MSE = .462$, $p < .01$. The four conditions were compared to each other using Bonferroni corrected pairwise comparisons. Recognition in the inter condition was better than in the extra condition ($p < .05$). Recognition in inter and extra conditions was better than in both ortho conditions ($p < .01$). Finally, recognition performance did not differ in the two ortho conditions ($p = .41$). These results are difficult to explain by approaches using alignment of a 3D representation [15] because such a differential effect of rotation direction would not be expected. Moreover, human performance questions the biological plausibility of the linear combination approach for face recognition [25], because it cannot explain why performance in the inter condition was better than in the extra condition. The results can for example be understood by a linear interpolation within an RBF network [19] – in the next section, we present another view-based framework, which can model the results [29]. Both of these models predict inter > extra > ortho, which was shown clearly in the psychophysical data. Interestingly, the results of the present study lead to the same conclusions as the study in [5], who used paper clips and amoeboid objects in order to investigate how humans encode, represent and recognize *unfamiliar* objects. In contrast, in our study perhaps the most familiar object class was used. Thus, familiarity with the object class does not necessarily predict qualitatively different viewpoint dependence.

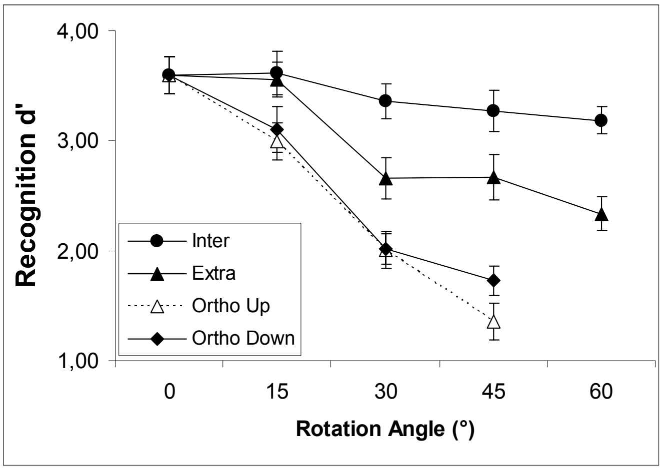

Figure 16 Human recognition performance in the four rotation conditions (inter, extra, ortho up, ortho down) across viewpoint (0° is the frontal view).

6.4 Experiment 11: Computational modelling of view-based performance

6.4.1 Description of the system

The original inter-extra-ortho experiment was analyzed using radial basis function (RBF) networks, which were able to capture the performance of subjects in the various tasks (see also [18] for a study on face recognition using RBF networks). In this paper, another kind of view-based computational model was used to explain the psychophysical data, which is based on a framework proposed in [29]. The motivation for the proposed framework came from several lines of research in psychophysics: First of all, evidence for a view-based object representation - as already stated above - has been found in numerous studies (also from physiological research). In addition, recent results from psychophysical studies showed that the temporal properties of the visual input play an important role for both learning and representing objects [28]. Finally, results from psychophysics (see e.g., [13,21,22] and previous chapters) support the view that human face recognition relies on encoding and storing local information contained in facial parts (featural information) as well as the spatial relations of these features (configural information).

A model, which can incorporate elements of these findings, was proposed in [29]. The framework is able to learn extensible view-based object representations from dynamic visual input on-line. In the following, the basic elements of the framework as used in this study are described. Each image is processed on multiple scales to *automatically* find interest points (in our case, corners). A set of visual features is constructed by taking the positions of the corners together with their surrounding pixel patches (see Figure 17). In order to match two sets of visual features, an algorithm based on [20] was used: It constructs a pair-wise similarity matrix **A** where each entry A_{ij} consists of two terms:

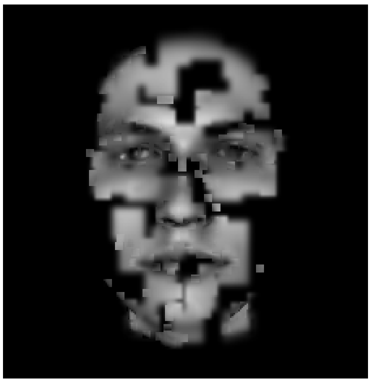

Figure 17 Feature representation as used by the computational framework – note that features focus on areas of interest (eyes, mouth, nose).

$$A_{ij} = \exp(-\frac{1}{\sigma^2_{dist}}\operatorname{dist}^2(i,j)) \cdot \exp(-\frac{1}{\sigma^2_{sim}}\operatorname{sim}^2(i,j)) \quad (1)$$

where $\operatorname{dist}^2(i,j) = ((x_i - x_j)^2 + (y_i - y_j)^2)$ measures the distance between a feature pair and sim(i,j) measures the pixel similarity of the pixel patches (in our case, using Normalized Cross Correlation). The parameters $\sigma_{dist}, \sigma_{sim}$ can be used to weight distance and pixel similarity. Based on the SVD of this matrix A=UVWT a new matrix is constructed a re-scaled, A'=UWT, which is then used to find a feature mapping between the two sets [20,29]. The goodness of the match is characterized by the percentage of matches between the two feature sets. This feature matching algorithm ensures that both global layout and local feature similarity are taken into account. It is important to note that there is neither a restriction to a global spatial measure in pixel space nor to a local measure of pixel similarity. Any kind of view-based feature measure can be introduced in a similar manner as an additional term in equation (1).

One of the advantages of this framework, which a purely view-based holistic representation lacks, is its *explicit* representation of local features. This enables the system amongst other things to be more robust under changes in illumination and occlusion [29,30]. Since the input consists of image sequences the visual features can also be augmented with *temporal information* such as trajectories of features. Temporal information is given in our case by the learning trials in which a small *horizontal* rotation is presented. We thus modified the distance term in equation 1 such that it penalizes deviations from the horizontal direction for feature matches by an increased weighting of the vertical distance between features *i* and *j*:

$$\text{dist}^2(i,j)) = ((x_i - x_j)^2 + \alpha(y_i - y_j)^2) \text{ with } \alpha \geq 1 \qquad (2)$$

Figure 18 shows matching features[13] between two images for two settings of α=1 and α=3: α=1 (Figure 18a) yields a matching score of 30 percent, whereas α=3 (Figure 18b) yields a matching score of 37 percent. The rationale behind using the penalty term not only comes from the dynamic information present in the learning phase, but is also motivated by the psychophysical results in [5,7], where humans showed a general tendency towards views lying on the horizontal axis.

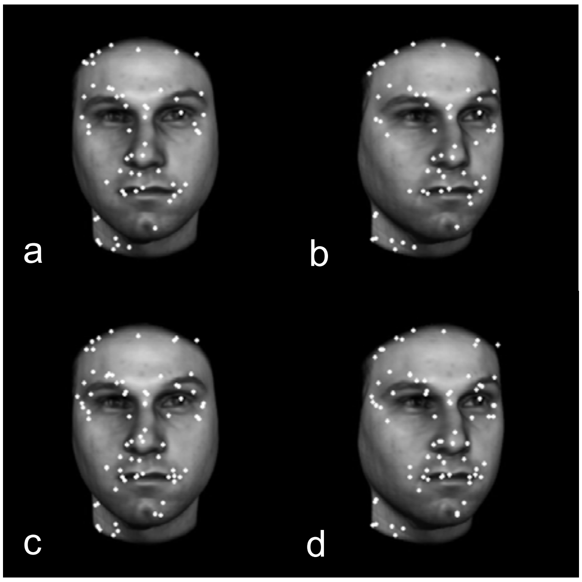

Figure 18 Matching features between two views of a face a) without vertical penalty term b) with vertical penalty term.

6.4.2 Computational recognition results

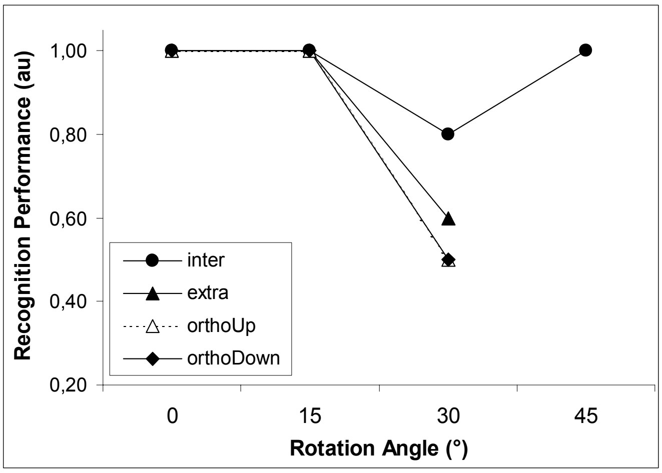

Figure 19 Machine recognition performance (arbitr. units) in the four rotation conditions (inter, extra, ortho up, ortho down) across viewpoint.

This section describes the computational recognition results, which were obtained using the same stimuli as for the human subjects. Again, the system was trained with two small image sequences around 0° and 60° and tested on the same views as humans. The final learned representation of the system consisted of the 0° and 60° view, each containing around 200 local visual features. Each testing view was matched against all learned views using the matching algorithm outlined above. To find matches, a winner-takes-all strategy was employed using the *combined matching score* of the two learned views for each face. The results in Figure 20 show that the computational

[13] Some matches between features are not exactly horizontal due to localization inaccuracies inherent in the corner extraction method.

model exhibits the *same* qualitative behavior in performance as human subjects replicating the drop in performance, i.e. inter>extra>ortho[14]. Inter performance was best due to support from two learned views as opposed to support only from the frontal view for the extra conditions. Recognition of ortho views was worst due to three factors: First, inter conditions had support from two views, second, the learned penalty term biased towards horizontal feature matches and third, the change in feature information for the same angular distance for faces is higher for vertical than for horizontal rotations. In Figure 20, inter and extra conditions are plotted only for view-changes up to 30°. For larger view changes a global correspondence cannot be established anymore since the available feature sets are too different. This observation agrees with findings from a study in [27]. In order to address the issue of generalization over larger view changes, the following extension to the framework is proposed, which consists of a two-step matching process: First, in order to determine head position the image is matched on a coarse scale against different views from a database of faces. This is possible since the global facial feature layout guarantees a good pose recovery even for novel faces [30]. The second step then consists of using more detailed part layout information to match parts consisting of groups of features to the image in the corresponding pose. Parts (which would correspond in the ideal case to facial parts such as eyes, mouth, nose, etc.) can again be matched using the same algorithm as outlined above under the constraint of the global part layout information. Such a constraint can easily be built into the matching process as a prior on the allowed feature deformations. Again, this proposed framework is consistent with evidence from psychophysical studies (e.g., [13,21,22 an previous chapters], see also [18] for a holistic two-stage model with alignment and view-interpolation).

In computational vision[15] the question how (facial) parts can be extracted from images and how a perceptually reasonable clustering of features can be created has recently begun to be addressed. A purely bottom-up way of extracting parts was suggested in [12], whereas [26,31] approach the issue from the perspective of categorization: extracting salient features, which maximize within-class similarity while minimizing between-class similarity. In [8], a 'Chorus of Fragments' was introduced, which is modeled after what/where receptive fields and also takes into account parts and their relations. One advantage of the framework proposed here, is its explicit use of features and their properties (such as pixel neighborhood, trajectory information, etc.), which provides the system with a rich representation and can be exploited for feature grouping. As shown in Figure 17, the visual features already tend to cluster around facial parts and in addition also capture small texture features of the skin (such as birthmarks and blemishes), which were hypothesized [27] to be important features for less view-dependent face recognition.

[14] The difference between the conditions was confirmed to be statistically significant by repeating the test 10 times with different sets of faces from the database.

[15] There is evidence from developmental studies that the basic schema of faces is innate [9,17], which could help newborn infants to learn encoding the parts of a face.

6.5 General Discussion

Several previous studies have investigated face processing under varying pose (for a short review and further results see [24]). In order to further understand the viewpoint dependent nature of face recognition it was investigated in this study whether qualitatively similar effects of viewpoint apply to face recognition as found in studies using unfamiliar objects like wire-frames and amoeboid objects [5,7]. Indeed, this was the case, the same qualitative effects of viewpoint were found, which were consistent with a view interpolation model of object recognition [5,18,19]. In addition, a computational model based on local features and their relations [29] showed the same qualitative behavior as humans. The breakdown of this model for large view-changes motivates an extension of the framework to explicitly model parts. At the same time, this framework should provide greater robustness against partial occlusion and less susceptibility to viewpoint changes due to the use of parts [10,27].
The concept of representing objects by their parts and spatial relations has been proposed many years ago by structural description theories (e.g., [1,16]). There are, however, several important differences between these approaches and the framework we propose here. First of all, in contrast to the traditional approaches by Marr and Biederman, it seems neither biologically plausible nor computationally possible to extract good edge-based representations as the input for recognition. Moreover, the parts proposed in this study are completely different both conceptually and computationally from the geons used in the approaches in [1,11]. Geons are defined by using Lowe's nonaccidental properties [14] and are meant to be viewpoint-independent (or at least for a certain range of views, see [2]). In contrast to [3], it proposed here that face–recognition relies on processing local part-based and configural information (which could also apply to many cases of object recognition). In contrast to geons, the parts are defined by grouping view-dependent image features. According to the type of features used, such parts are more or less viewpoint-dependent. Current experiments explore to what extent *part-based* representations in human face recognition are viewpoint dependent. RBC theory assumes that a small set of geons suffices to explain the relevant aspects of human object recognition. However, in many cases of everyday object recognition, defining the features is rather a matter of perceptual learning [23], and it seems more reasonable to believe that the number of parts represented by the human brain for recognition exceeds a 24 or 36 geon set by far and in addition might be heavily task-dependent.

6.6 References

1. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94(2), 115-147.
2. Biederman, I., Gerhardstein, P.C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. Journal of Experimental Psychology: Human Perception and Performance, 19, 6, 1162-1182.
3. Biederman, I., Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society London, B, 352, 1203-1219.
4. Blanz, V. , Vetter, T. (1999). A Morphable Model for the Synthesis of 3D Faces. In Proc. Siggraph99, pp. 187-194.

5. Bülthoff, H.H., Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. PNAS USA, 89, 60-64.
6. Collishaw, S.M., Hole G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29, 893-910.
7. Edelman, S., Bülthoff, H, H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Research, 32(12), 2385-4000.
8. Edelman, S. Intrator, N. (2000). A productive, systematic framework for the representation of visual structure. In Proc. NIPS 2000, 10-16.
9. Goren, C., Sarty, M., Wu, P. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics, 56, 544-549.
10. Heisele, B., Serre, T., Pontil, M., Vetter, T., and Poggio, T. (2001). Categorization by learning and combining object parts. In Proc. NIPS 2001.
11. Hummel, J.E., Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. Psychological Review, 99(3), 480-517.
12. Lee, D., Seung S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401:788-791.
13. Leder, H., Candrian, G., Huber, O., Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30, 73-83.
14. Lowe, D.G. (1985). Perceptual organization and visual recognition. Boston: Kluwer Academic Publishing.
15. Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence, 31, 355-395.
16. Marr, D. (1982). Vision. San Francisco: Freeman.
17. Morton, J., Johnson, M.H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. Psychological Review, 98, 164-181.
18. O'Toole, A. J., Edelman, S., Bülthoff H.H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. Vision Research, 38 , 2351- 2363.
19. Poggio T, Edelman S. (1990) A network that learns to recognize three-dimensional objects. Nature, 18, 343(6255), 263-266.
20. Pilu, M. (1997). A direct method for stereo correspondence based on singular value decomposition, In Proc. CVPR'97, 261-266.
21. Schwaninger, A., Mast, F. (1999). Why is face recognition so orientation-sensitive? Psychophysical evidence for an integrative model. Perception, 28 (Suppl.), 116.
22. Sergent J. (1985). Influence of task and input factors on hemispheric involvement in face processing. Journal of Experimental Psychology: Human Perception and Performance, 11(6), 846-61.
23. Schyns, P. G., Rodet, L. (1997) Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 681-696.
24. Troje, N. F., Bülthoff, H.H. (1996). Face recognition under varying pose: the role of texture and shape. Vision Research, 36, 1761-1771.
25. Ullman, S., Basri, R. (1991). Recognition by linear combinations of models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(10), 992-1006.
26. Ullman, S., Sali, E. (2000). Object Classification Using a Fragment-Based Representation. In Proc. BMCV'00, 73-87.
27. Valentin, D., Abdi, H., Edelman, B. (1999). From rotation to disfiguration: Testing a dual-strategy model for recognition of faces across view angles. Perception, 28, 817-824.
28. Wallis, G. M. , Bülthoff, H.H. (2001). Effect of temporal association on recognition memory. PNAS USA, 98, 4800-4804.
29. Wallraven, C., Bülthoff, H.H. (2001). Automatic acquisition of exemplar-based representations for recognition from image sequences. CVPR 2001 - Workshop on Models vs. Exemplars.

30. Wallraven, C., Bülthoff, H.H. (2001). View-based recognition under illumination changes using local features. CVPR 2001 - Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation.
31. Weber M., Welling M. and Perona P. (2000). Unsupervised Learning of Models for Recognition. In Proc. ECCV2000.

7 Learning from humans: computational modeling of face recognition

7.1 Abstract

In this paper a computational architecture of face recognition is proposed that is based on evidence from cognitive research. Using an implementation of this architecture it was possible to model aspects of human performance, which were found in psychophysical studies. In addition, the computational results gave rise to a number of predictions, which in turn lead to further psychophysical experiments. Thus, this study is an example of closing the loop between psychophysics and computational modelling in order to achieve a deeper understanding of the complex cognitive processes underlying face recognition in the human brain.

7.2 Introduction

Faces are one of the most relevant stimulus classes in everyday life. Although they form a very homogenous stimulus class, adult observers are able to detect subtle differences between facial parts and their spatial relationship. According to Bahrick et al. [1975] we are able to recognize familiar faces with an accuracy of 90 per cent or more, even when some of those faces have not been seen for fifty years. Moreover, whenever people interact facial expressions are automatically interpreted in order to identify underlying emotional states. These evolutionary very adaptive abilities seem to be remarkably disrupted if faces are turned upside-down.

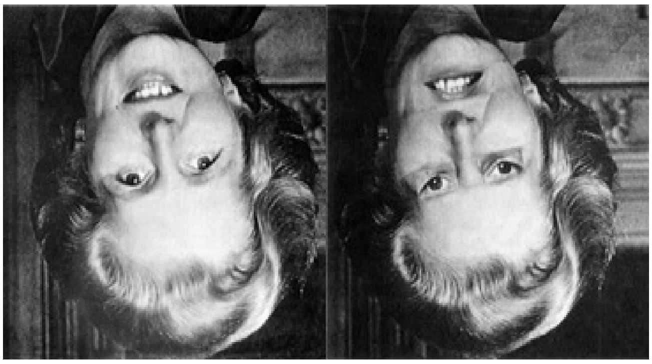

Figure 20. Thatcher illusion. When the pictures are viewed right side up, the face on the right appears highly grotesque. This strange expression is much less evident when the faces are turned upside-down (as above [Thompson, 1980])

Consider the two pictures in Figure 20: Recognizing the depicted identity is more difficult when faces are inverted. Moreover, the two faces seem to have a similar facial expression. Interestingly, if the two pictures are turned right side up, one can easily identify the depicted person and grotesque differences in the facial expression are revealed [Thompson, 1980]. As pointed out by Rock [1973] rotated faces seem to overtax an orientation normalization mechanism such that it is not possible to perceive mentally rotated faces as wholes. Instead, rotated faces seem to be processed by matching parts, which could be the reason why in Figure 20 the faces look normal when turned upside-down.

Young et al. [1987] discovered another interesting effect. They created composite faces by combining the top and bottom half of different faces (Figure 21). If the two halves were aligned and presented upright, a new face resembling each of the two originals seemed to emerge. This made it very difficult to identify the persons from either half. If the top and bottom halves were horizontally misaligned, then the two halves did not spontaneously fuse to create a new face, and the constituent halves remained identifiable. Interestingly, when these stimuli were inverted (turn Figure 21 around), the constituent halves of the aligned and misaligned displays were equally identifiable. Furthermore, the subjects were significantly faster at naming the constituent halves in inverted than in upright composites. The authors interpret this result as "a dramatic illustration of the absence of interference from configurational information in the inverted composites" (p. 753, [Young et al., 1987]). When upright, the alignment of face composites creates a new configuration which resembles a new face. When inverted the processing of configural information seems to be impaired and the two identities are easier to extract based on the facial parts alone.

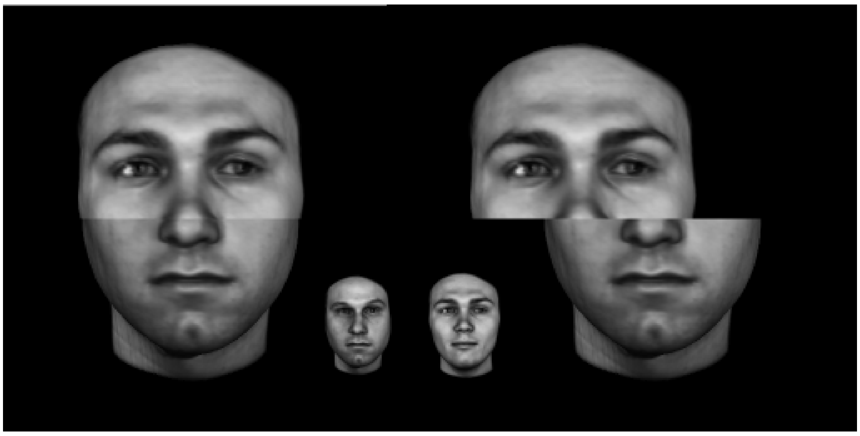

Figure 21. Aligned and misaligned halves of different identities. A new identity seems to emerge from the aligned composites (left), which makes it more difficult to extract the original identities. This does not occur for the misaligned composite face (right). It also does not occur for *both* images if you turn the images by 180 degrees. The insets show the constituent faces.

7.3 Components and configural processing in face recognition

The distinction between parts or component information on one hand and configural information on the other has been used by many studies on human face recognition (for an overview see [Schwaninger et al., in press]). The term component information (or featural, piecemeal, part-based information) has been referred to facial elements, which are perceived as distinct parts of the whole such as the eyes, mouth, nose or chin [Carey and Diamond, 1977; Sergent, 1984]. In contrast, the term configural information refers to the spatial relationship between components [Bruce, 1988] and has been used for distances between parts (e.g. inter-eye distance or eye mouth distance) as well as their relative orientation.

There are several lines of evidence in favour of the assumption of component vs. configural representations in face processing. One of the first demonstrations for qualitative differences has been provided by Sergent [1984]. She used pairs of faces that were mismatched either in the eyes or face contour (change of component information) or in the internal spacing of features (change of configural information). The analysis of her results revealed that for upright faces configural and component information was used. In contrast, no evidence was found for the use of configural information in upside-down faces. Note that Sergent [1984] used highly schematic faces that could make it difficult to generalize from this result to the processing of real faces. However, Searcy & Bartlett [1996] found comparable results for colored photographs using different experimental methods. Again, their results suggested that in upright faces component and configural information is used, whereas in inverted faces the processing of configural information is hampered. Another demonstration of the differential effects of orientation upon processing component and configural information was found in [Schwaninger and Mast, 1999]. In their experiment two faces were presented sequentially and the subjects had to detect whether components were changed (eyes and mouth replaced) or whether configural information was altered (increased inter-eye distance and distance between the eyes and mouth). Whereas both types of alterations were easy to detect in upright faces, the participants had a hard time in detecting configural changes when the faces were rotated. In contrast, the detection of component changes was almost unaffected by rotation.

In short, these and other studies provide converging evidence in favour of a *qualitative distinction* between component and configural information in face processing. However, one possible caveat of studies that investigated the processing of component and configural information by replacing or altering facial parts is the fact that such manipulations are difficult to conduct selectively. Replacing the nose (component change) sometimes alters the distance between the contours of the nose and the mouth and might changes the configural information. Similar difficulties apply to configural manipulations when they are conducted by changing the relative position of components. For example moving the eyes apart (configural change) can lead to an increase of the bridge of the nose, which is a component change.

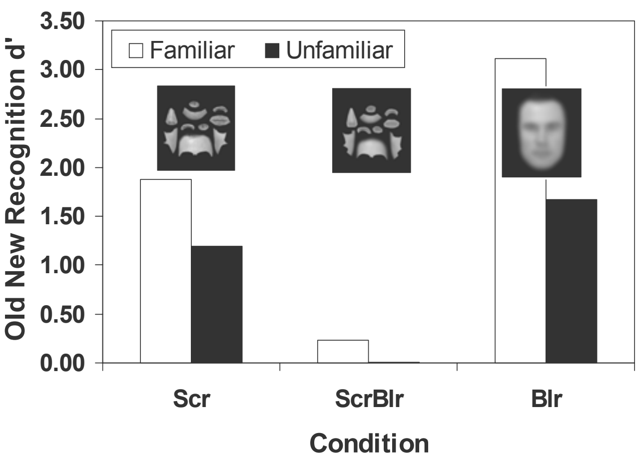

Figure 22. Results from the study described in chapter 4. Previously learnt intact faces could still be recognized when they were scrambled into their components (Scr). Applying a low pass filter made the scrambled version impossible to recognize (ScrBlr). When the same filter was applied to whole faces recognition of these configural versions was well above chance (Blr). When the target faces were familiar (white bars), recognition performance increased but the relative balance between component and configural recognition remained the same.

Problems like these were avoided in the psychophysical study described in chapter 4. In contrast to previous studies, a method was used that did not alter configural or featural information, but eliminated either the one or the other. The results of two experiments are again depicted in Figure 22, where the recognition performance is measured in d'-scores, which is a common measure in psychophysics. D' is related to ROC-analysis and is defined as the z-transform of the *hit rate* (percentage of correctly identified test images) minus the z-transform of the *false alarm rate* (percentage of images, which were erroneously identified as learnt images).

In Experiment 7 it was found that previously learnt faces could be recognized by human participants even when the faces were scrambled into constituent parts (or components) so that configural information was eliminated (Figure 22, left). This result is consistent with the assumption of *explicit representations* of component information in visual memory. In a second condition, a low pass filter that made the scrambled part versions impossible to recognize was determined (Figure 22, middle). This filter was then applied to whole faces in order to create stimuli in which by definition local part-based information is eliminated and it can be tested whether configural information is explicitly encoded and stored. It was shown that configural versions of previously learnt faces could be recognized reliably (Figure 22, right), suggesting separate explicit representations of configural information. In Experiment 8 these results were replicated for subjects who knew the target faces (white bars in Figure 22). Both Experiments provided converging evidence in favor of the view that recognition of familiar and unfamiliar faces relies on component and configural information.

7.4 An integrative model of face processing

Based on the results from different psychophysical studies described in previous chapters the model depicted in Figure 23 has been proposed, which summarizes the cognitive architecture of face recognition. Processing entails extracting *local part-based information* and global *configural relations* between them in order to activate component and configural representations in higher visual areas (so-called face selective areas). In the study described in chapter 5, repetition priming[16] was used in order to investigate whether the outputs of component and configural representations converge to the same face identification units. Since priming was found from scrambled to blurred faces and vice versa it was proposed that the outputs of component and configural representations converge to the same face identification units.

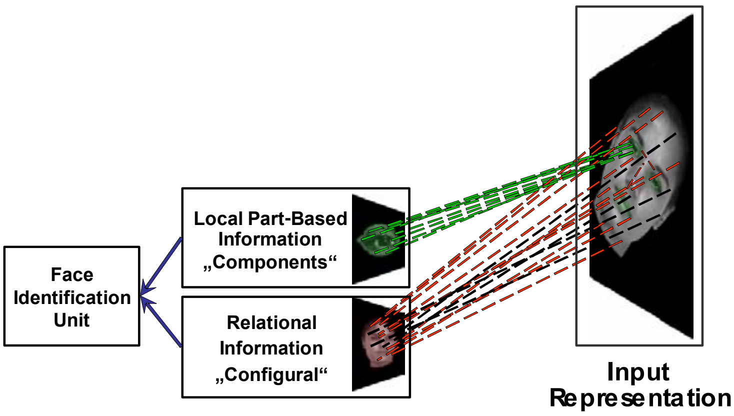

Figure 23. Integrative model for face recognition.

7.5 Computational modeling

Based on the psychophysical experiments outlined above, a computational architecture has been constructed, which captures the key findings of the model in Figure 23. An implementation of this architecture is then used to simulate the same experiments in order to compare human with machine performance.

[16] Priming in psychophysical experiments means that once you have seen a stimulus before, you are faster to recognize it in subsequent tests. If one can find priming for *different* stimuli, it generally is taken to be evidence for a common representation of the stimuli in the brain.

7.5.1 Stimulus generation

Using the MPI face database [17], it was possible to construct the experimental stimuli from the previous sections in order to use them as input to the computational system. This face database was constructed from high resolution three-dimensional laser-scans of 200 individuals. Each scan consists of ~70000 vertices which contain both x,y,z and R,G,B coordinates. A post-processing [Blanz and Vetter 1999] guarantees that all faces are in direct correspondence such that, e.g., the tip of the nose can be robustly located in all faces. This property was used to reproduce the set of stimuli used in the psychophysical experiments with a set of 20 male faces. Each face was thus rendered as a 512x512 pixel image in a neutral pose under frontal illumination in 4 conditions: full, blurred, scrambled and scrambled-blurred. A standard Gaussian blur filter of size 5x5 pixels was used, which was applied several times to produce a total of 5 blur levels.

7.5.2 The face representation

One of the most important elements of the psychophysical model depicted in Figure 23 is an explicit separation of configural and component information within the object representation. The computational implementation is motivated by studies, where a first version was successfully used to model psychophysical results on view-based object recognition [Wallraven and Bülthoff 2001a;b; Wallraven et al. 2002].

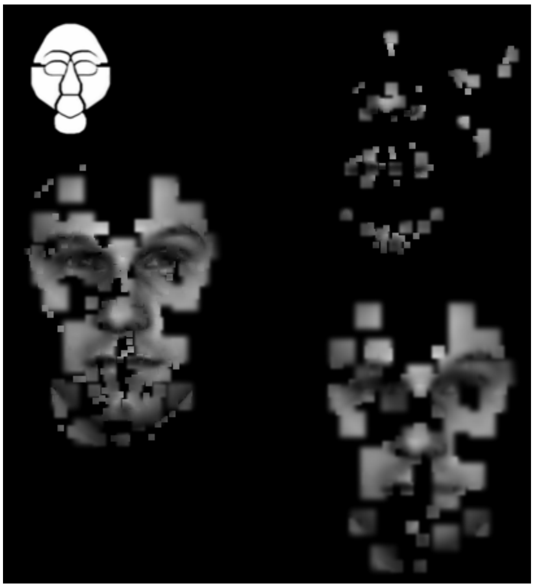

Figure 24. A reconstruction of a full face (left) from its configural (bottom right) and component (top right, showing only nose, mouth, chin and right cheek) features. The inset shows the mask used to define the face components.

The algorithm for constructing the face representation proceeds as follows: First, an input image is processed on three scales of a Gaussian pyramid to extract visual features. A visual feature is found by using an interest point detector (such as a standard corner detector [Wallraven and Bülthoff 2001a]), which yields pixel coordinates of salient image regions. In addition to this positional information, the pixel values of a small neighborhood of 5x5 pixels are included as appearance information. This approach is motivated from both psychological and physiological studies, which support the notion of visual features of intermediate complexity in higher brain areas (for an overview, see, e.g., [Ullman et al. 2002]).

The configural part of the representation is then formed by the features from the low resolution scales (Figure 24, bottom right), which corresponds to a coarse face template capturing a few highly salient facial features. The finer scales are used to construct the components of the face (Figure 24, top right), thus capturing detailed information about the face components. One of the important steps here is to detect which features belong to which component of the face. In the current implementation, this step is solved by

[17] A part of this database is freely available at `http://faces.kyb.tuebingen.mpg.de`

prior knowledge from the face database: as each face is in correspondence, one face can be used to identify the components. This information is then used to label each pixel in the image as belonging to, e.g., the forehead or the left eye (see Figure 24, left). This form of supervised learning was used as there is evidence from developmental studies that humans possess an innate ability to detect important facial features (i.e., eyes, mouth, nose)[18].

7.5.3 Recognition of face images

The algorithm for recognition of new test images is the second main part of the computational modelling. As each image consists of a set of visual features, recognition in our case amounts to finding the best matching feature set between a test image and all training images. As the psychophysical experiments point towards two routes for face processing, two types of matching algorithms for recognition based on configural and component information were implemented.

Matching of two feature sets is done by an algorithm described in [Pilu 1997]. First, a similarity matrix **A** is constructed between the two sets, where each term A_{ij} in the matrix is of the form:

$$A_{ij} = \exp(-\frac{1}{\sigma_{dist}^2}\mathrm{dist}^2(i,j)) \cdot \exp(-\frac{1}{\sigma_{sim}^2}\mathrm{sim}^2(i,j)) \qquad (1)$$

The first term in equation (1) specifies the positional similarity (i.e., the pixel distance in the image), whereas the second term measures the appearance similarity (i.e., the cross-correlation of the two pixel patches) of two visual features. The parameters σ_{dist} and σ_{sim} can be used to control the relative importance of the two types of information. The matrix A thus captures similarity between two feature sets based on both distance information and appearance information. Corresponding features can now be found by looking at the largest elements of **A** both in row and column [Pilu 1997; Wallraven & Bülthoff 2001a;b], which yields a one-to-one mapping of one feature set onto the other.

In the case of *configural matching* the Singular Value Decomposition (SVD) of **A** is employed to find a set of one-to-one matches between two images. As this approach makes use of the full matrix (~100 low level features in the configural representation) it represents a *global* (configural) matching strategy with stronger emphasis on the positional similarity of features.

Matching of components on the other hand relies more heavily on appearance and usually involves only a few features. The scrambled-blur experiment involves two conditions for component matching: In the first condition, full information from the facial feature detectors is available, thus reducing the search for matching features to the correct components. In this case, the same matrix **A** is used, which is constructed for each pair of components. A simple greedy search strategy is then used to find the biggest elements both in row and column. In the case of scrambled or heavily blurred images however, it might not be possible to extract reliable locations of parts. In this case, matching would amount to finding the correct correspondences of a small set of features within the full set of features. To this end, a simple exhaustive search paradigm was used, which tries to find the best pixel position maximizing the number of matches for each component. One might see this strategy as an implementation of an attentional

[18] There are, however, several recent face detectors, which perform quite well in detection of face components, see, e.g., [Hjelmas and Low, 2001].

focus scanning over the image, where at each location the output of the component detectors is recorded. In the following section, both conditions will not explicitly be separated as they produced similar results.

7.5.4 Modeling results

The set of 20 faces was divided into a training set of 10 faces and a distractor set of 10 faces. Training was done by extracting visual features of the 10 whole faces from the training set. For testing, the system was first presented with a number of scrambled stimuli with increasing blur-level. The percentage of matches between the test image and all trained whole faces was evaluated with the best matching face being determined by the maximum percentage. This was done separately for both the configural and the component route. Figure 25 shows the results from this first experiment again using d'-scores to facilitate comparison with Figure 22. D'-scores drop with increasing amount of blur for both configural and component matching. Component matching performs worse than configural matching in a similar way as in the psychophysical experiments. This is mainly due to the fact that each component contains only a few features which lose their high selectivity with increasing amount of blur. Another interesting finding is that some of the components have a higher saliency than others (shown in Figure 25 as "salient comp."). These components are eyes, mouth, nose and chin, which is not surprising since these are also the components, which come most readily to mind when thinking about faces. In a more specific way, however, this difference can be attributed to the visual features of the component representation. Each component consists of a small number of detailed features, which were defined as "interest" points (i.e., points with a high curvature in pixel intensity space). The average saliency of the features within the mouth region, however, is much higher than the average saliency of features in the cheek region as the amount of facial texture variation is higher in the former than in the latter. This

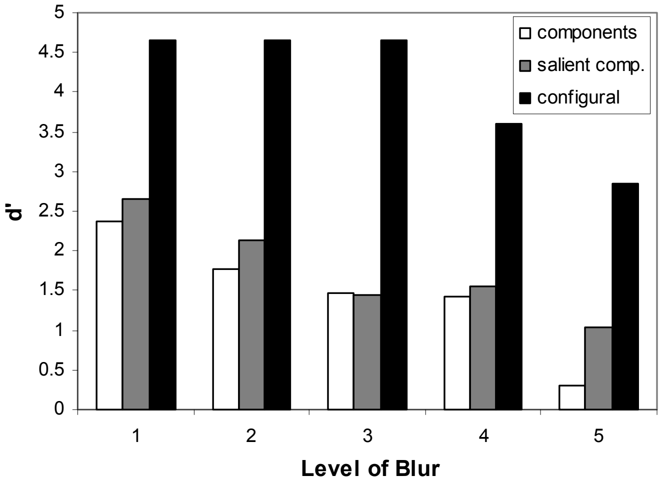

Figure 25. d'-values for different blur-levels 1-5 for component and configural matching.

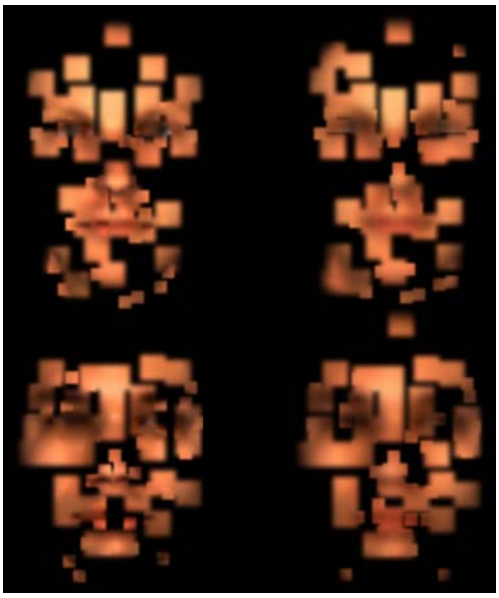

Figure 26. Two faces in the configural representation for the original condition (left column) and for the most blurred condition (right column). Note that the global layout of the features stays roughly the same despite the large change in blur level.

fact can then explain why there is a difference between the components with regard to the amount of blurring which is needed to disrupt their information content.
Configural information on the other hand consists of a number of visual features from low resolution scales, which are matched globally. Thus, this route is more resistant to changes in blur level as the global layout of features in low resolutions stays roughly the same (see Figure 26).

7.6 Conclusion

Psychophysical evidence strongly supports the notion that face processing relies on two different routes, which are represented by configural information and component information. In this study a computational model of such a processing architecture was implemented, which is based on the explicit separation of the two types of information. Visual features consisting of appearance and position information are at the basis of the proposed representation. The configural route is implemented by two key-elements. The first element concerns the representation itself, which is formed in this case by visual features from a low resolution scale. The second important element is the process by which this representation is matched against other inputs: An algorithm, which implements a global matching of the visual features. The representation of the component route is formed by the visual features from the detailed resolution scales, whereas the processing of this representation is done with a simple local matching algorithm. This distinction between representations and processes is very important if one wants to come to a full understanding of the cognitive processes underlying face (and, indeed, object) recognition. Furthermore, it seems to be the case by looking at the psychophysical data that global processing is very orientation-sensitive whereas local processing is not. It is the aim of current modelling efforts to extend the computational architecture in order to model the two types of illusions mentioned in the introduction.
In a second step, the architecture was tested using the same stimuli as in the psychophysical experiments in order to examine its performance and modelling capabilities. The results were very similar on a qualitative level to human performance. In this context it has to be said that an exact quantitative modelling – while this might seem a desirable goal – cannot be realistically achieved as there are too many hidden variables in the exact formation of the psychophysical data. A qualitative similarity on the other hand is a sign that the basic assumptions of the computational architecture and its implementation share a similar structure. This argumentation leads to the second important step in biological cybernetics: Closing the loop between psychophysics and computational modelling. For example the computational experiments have shown that certain features seem to have a higher saliency than others. In this case, a psychophysical experiment which validates these findings will give even stronger support for the proposed architecture.

7.7 References

[Bahrick et al., 1975] H.P. Bahrick, P.O. Bahrick and R.P. Wittlinger. Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General,* 104, pages 54-75.

[Blanz and Vetter 1999] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *Proc. Siggraph99*, pp. 187-194.

[Bruce 1988] V. Bruce. *Recognising faces*. Hillsdale, NJ: Laurence Erlbaum.

[Carey and Diamond, 1977] S. Carey and R. Diamond. From piecemal to configurational representation of faces. *Science,* 195(4275), pages 312-314.

[Hjelmas and Low, 2001]. E. Hjelmas and B.K. Low. Face Detection: A Survey. *Computer Vision and Image Understanding*. 83. pages 236-274.

[Pilu 1997] M. Pilu. A direct method for stereo correspondence based on singular value decomposition, In *Proc. Computer Vision and Pattern Recognition 1997*, pages 261-266.

[Rock 1973] I. Rock. *Orientation and form*. New York: Academic Press.

[Schwaninger et al., 2002a] A. Schwaninger, J. Lobmaier, and S.M. Collishaw. Role of featural and configural information in familiar and unfamiliar face recognition. In *Proceedings Biologically Motivated Computer Vision*, pages 643-650, Tuebingen, Germany, November 2002.

[Schwaninger et al., in press] A. Schwaninger, C.C. Carbon and H. Leder. Expert face processing: Specialisation and constraints. In G. Schwarzer and H. Leder, *Development of face processing*.

[Schwaninger et al., 2002b] A. Schwaninger, S.M. Collishaw and J. Lobmaier. Role and interaction of featural and configural processing in face recognition. Vision Sciences Society, second Annual Meeting, May 10 - 15, Sarasota, Florida, USA.

[Schwaninger and Mast, 1999] A. Schwaninger and F. Mast. Why is face recognition so orientation-sensitive? Psychophysical evidence for an integrative model. *Perception (Suppl.)*, 28, page 116.

[Searcy and Bartlett, 1996] J.H. Searcy and J.C. Bartlett. Inversion and processing of component and spatial-relational information of faces. *Journal of Experimental Psychology: Human Perception and Performance,* 22(4), pages 904-915.

[Sergent 1984] J. Sergent. An investigation into component and configurational processes underlying face recognition. *British Journal of Psychology,* 75(Pt2), pages 221-242.

[Thompson 1980] P. Thompson. Margaret Thatcher -- A new illusion. *Perception*, 9(4), pages 483-484.

[Ullman et al., 2002] S. Ullman, M. Vidal-Naquet and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), pages 682-687.

[Wallraven and Bülthoff, 2001a] C. Wallraven, H.H. Bülthoff. Automatic acquisition of exemplar-based representations for recognition from image sequences. CVPR 2001 - Workshop on Models vs. Exemplars.

[Wallraven and Bülthoff, 2001b] C. Wallraven, H.H. Bülthoff. View-based recognition under illumination changes using local features. CVPR 2001 - Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation.

[Wallraven et al., 2002] C. Wallraven, A. Schwaninger, S. Schuhmacher and H.H. Bülthoff. View-Based Recognition of Faces in Man and Machine: Re-visiting Inter-Extra-Ortho. In *Proceedings Biologically Motivated Computer Vision*, pages 651--660, Tuebingen, Germany, November 2002.

[Young et al.., 1987] A.W. Young, D.J. Hellawell and D.C. Hay. Configural information in face perception. *Perception,* 16(6), pages 747-759.

8 Expert face processing: Specialization and constraints

8.1 Introduction

Face processing in adults is the product of innate mechanisms, and is also based on years of experience. There is no doubt that face processing is a human skill at which most adults are real experts. In the this last chapter theories and hypotheses are reviewed concerning adults' face processing skills, as well as what information and processes these are based on. The results from previous chapters are integrated and compared to other findings. Moreover, it is discussed how the high specialization is attained at the cost of being susceptible to specific conditions and how that can be explained with the integrative model for face recognition that was presented in previous chapters.

Expertise, according to the American Heritage Dictionary is given when a person shows a high degree of skill in or knowledge of a certain subject. This definition implies that an expert is a high-grade *specialist*. Expertise does not have to be accessible in an explicit way, because an expert does not have to know all the facts of his expertise. The skill humans show in identifying faces is astonishing. As mentioned earlier, according to Bahrick, Bahrick, and Wittlinger (1975) adults are able to recognize familiar faces with an accuracy of 90 per cent or more, even when some of those faces have not been seen for fifty years. Moreover, faces are a class of objects which encourage a special kind of categorization. According to the logic of Roger Brown's seminal paper "How shall a thing be called" (Brown, 1958), the level of the object name reflects the entry point of the recognition process. When asked to name pictures of faces spontaneously, humans produce the concrete names of the persons shown.

Classifying objects at this kind of subordinate level is typical of experts (Tanaka & Taylor, 1991). Expertise can not only be recognized by the frequency of subordinate-level classifications but also by the speed of word generation (Tanaka, 2001a): Adults identified faces as fast at the subordinate level (the name of the person) as at the basic level (e.g., "human"). This is clear evidence for a level of expertise.

To understand the development of face processing from childhood to adulthood better, it is necessary to review the characteristics of information processing used by adults. First, different types of pictorial information contained in faces are discussed. Then the holistic hypothesis and the schema hypothesis are reviewed. This is followed by a discussion of important characteristics of adult face recognition, namely the sensitivity to configural information and the specialization in upright faces. Subsequently, the component configural hypothesis is discussed. Finally, the integrated model of face recognition established in previous chapters is applied to understand the main aspects of a fully developed face processing system.

8.2 Information contained in faces

Faces are complex three-dimensional surfaces of the front side of the human head. Psychophysical studies using computer graphics have distinguished surface-based

shape information from superficial properties such as color and texture (e.g., Hill, Schyns, & Akamatsu, 1997; Troje & Bülthoff, 1996).
Another commonly used distinction is based more on phenomenology. The term *component* information (or componential, piecemeal, featural information) has been used to refer to separable local elements, which are perceived as distinct parts of the whole such as the eyes, mouth, nose or chin (Carey & Diamond, 1977; Sergent, 1984). Components describe the basic primitives in faces, and the number of dimensions on which all components can differ provides the basis for all human faces being unique. A second type of information has been referred to as *configural* or *relational*. According to Bruce (1988), the term configural information refers to the "spatial interrelationship of facial features" (p. 38), i.e., features which come about from spatial arrangements, such as eye-distances, nose-mouth-distance. Distinctiveness correlates positively with the recognizability of faces, and Leder and Bruce (1998) revealed that component as well as configural information contribute to the distinctiveness of faces. Configural information was defined further by Diamond and Carey (1986). They used the term *first-order relational information* for the basic arrangement of the parts and *second-order relational information* to refer to specific metric relations between features.
The term *holistic* has been used to describe representations that store a face as an unparsed perceptual whole without specifying the parts explicitly. It has been operationalized in whole-to-part-superiorities and refers to properties and features when the face is processed as a Gestalt and not parsed into components (Farah, Tanaka, & Drain, 1995; Tanaka & Farah, 1993). A simple two-dimensional analogy for a holistic face representation would be a bitmap that only specifies the color values of points without providing any information about which points belong to the mouth or the eyes. Although the bitmap contains eyes and a mouth, it does not represent them explicitly[19].
These different types of information contained in faces are related to hypotheses about adult face processing which are discussed next.

8.3 Mechanisms of face processing in adults

In order to explain the mechanisms used in adult face processing, several hypotheses have been proposed. According to the holistic hypothesis, adults process faces as unparsed perceptual wholes. The schema hypothesis assumes that the ability to process faces improves over many years and is attained at the expense of flexibility. This specialization could be related to adults' high sensitivity to configural information. Since faces are usually seen upright, it is not surprising that orientation is a critical variable for a face processing system that develops from years of experience. According to the component configural hypothesis, the processing of configural information is much more impaired by changes of orientation than the processing of component information. Why this might be the case is explained by the integrative model we propose after discussing each of these hypotheses in more detail.

[19] Note that this definition is different from the concept of holistic processing, which is understood in terms of overall similarity relations (see Chapter 4).

8.4 Holistic hypothesis

According to the holistic hypothesis, upright faces are stored as unparsed perceptual wholes in which individual parts (components) are not explicitly represented (Farah et al., 1995; Tanaka & Farah, 1993). Several empirical findings have been interpreted in favor of this view. For example, Tanaka and Farah (1993) reasoned that if face recognition relies on parsed representations, then a component (e.g., a nose) presented in isolation should be easy to recognize. In contrast, if faces are represented as unparsed perceptual wholes (i.e., holistically) then a part of a face presented in isolation should be much more difficult to recognize. In their experiments, participants were trained to recognize upright faces, each of which had a different pair of eyes, nose, and mouth. In the test phase, images of faces were presented in pairs. Each pair of faces differed only in the shape of one part of the face. In one test condition, two facial parts were presented in isolation. The subjects had to judge which of the two parts belonged to a face familiar from the training phase. In the whole face condition, the parts were embedded in the facial context. For example, one face contained the original nose and the other contained a different nose. The participants had to judge which of them was the face familiar to them from the training phase. Parts presented in isolation were more difficult to identify than whole faces. In contrast, when participants were trained to recognize inverted faces, scrambled faces, and houses no advantage of presenting the parts in their context was found. The authors interpreted this result in favor of the holistic hypothesis and proposed that face recognition relies mainly on holistic representations while the recognition of objects is based much more on part-based representations. Whereas encoding and matching parts are assumed to be relatively orientation-invariant (Biederman, 1987), holistic processing is thought to be very sensitive to orientation (see also Biederman & Kalocsai, 1997; Farah et al., 1995).

The results of a study conducted by Tanaka and Sengco (1997) provide further support for the holistic hypothesis, although their concept of holistic is slightly different. Instead of assuming that faces are processed as unparsed perceptual wholes, the authors reasoned that if *both component and configural information are combined* into a single holistic representation, changes in configural information should affect the recognition of facial parts (component information). This was precisely what was found in their first experiment: After training with upright faces, the subjects recognized components (eyes, nose and mouth) better in the unaltered facial context than in the context of a face in which the configural information had been changed by manipulating the distance between the eyes. If holistic processing is hampered by inversion and if face recognition relies much more on holistic representations than object recognition does, then a similar configural manipulation should have no effect on the recognition of parts of inverted faces or objects such as houses. This indeed was the case. The authors showed that configural manipulations did not affect the recognition of isolated parts when faces were presented upside-down nor did they do so when upright houses were used in the training and test conditions. (For faces, the alteration of configural information was accomplished by increasing the distance between the eyes, and for houses by manipulating the distance between the windows.) Thus, altering the configural information only affects the recognition of parts in the case of upright faces. This finding favors the view that in normal (upright) face processing the component and configural information is combined into a single holistic representation and that this holistic processing is disrupted by inversion.

Another line of evidence for this view is derived from a study carried out by Rhodes, Brake, and Atkinson (1993). These authors used (coarse) digitized versions of full-face photos in a recognition memory paradigm. Configural alterations, which were induced by altering the internal spacing of the eyes and mouth, were more difficult to recognize when faces were inverted. Interestingly, when the eyes or mouth were replaced with those of another face, effects of inversion were even more detrimental to recognition performance! Rhodes et al. (1993) concluded that either the component changes also affected the configural information or that the assumption that component processing is relatively unaffected by inversion is incorrect. The authors reasoned that if the replacement of components also resulted in a configural change and this caused the decrease in performance for inverted faces, then this effect of inversion should disappear when the components are presented alone. The results of their Experiment 2 favoured this interpretation. In line with the results of Tanaka and Sengco (1997), the findings of Rhodes et al. (1993) are consistent with the view that in normal (upright) face processing component and configural information is combined into a single holistic face representation and that this holistic processing is impaired by inversion. Note that this concept of holistic processing differs slightly from the original definition of Tanaka and Farah (1993) and Farah et al. (1995). In the original view, holistic processing just means that parts are not represented explicitly. In contrast, holistic processing according to the results of Tanaka and Sengco (1997) and Rhodes et al. (1993) would imply that component and configural information are first encoded separately and then integrated into a holistic representation.

According to Farah et al. (1995) the holistic hypothesis also predicts that effects of inversion can be eliminated if participants are induced to represent faces in terms of their parts. Indeed, these authors found that inversion had the expected negative effect on the recognition of faces that were studied normally, while this impairment disappeared when faces were studied as parts (head outline, eyes, nose, and mouth presented simultaneously in different boxes). However, while the authors admit that it is possible to represent faces in terms of their components, they stress that performance is impaired by inversion because faces are usually represented holistically, i.e., parts are not represented explicitly.

An alternative definition of holistic processing of faces was tested by Macho and Leder (1998). Holistic processing could be achieved by an interactive feature processing in which the processing of one feature depends in general on the quality of another feature. In a similarity decision task using faces which systematically varied on two or three dimensions to target faces, they did not find evidence for this kind of interactive processing.

8.5 Schema hypothesis

Goldstein and Chance (1980) have suggested another hypothesis. According to their view, the ability to process faces (i.e., the face schema) improves with exposure to them. These authors suggest that this improvement is attained at the expense of flexibility. Therefore, because faces are usually seen upright, it follows that recognition performance should improve with age, but performance with unusual stimuli such as inverted faces should decline through development. Their predictions have been supported by studies that investigated the development of face recognition (for reviews see Carey, 1992; Ellis, 1992; Johnston & Ellis, 1995). A study by Diamond and Carey (1986) provides another line of evidence in favour of the schema hypothesis. These authors used faces and dog profiles as stimuli. They found that the performance of novices was affected by inversion when tested with human faces but not when dog profiles had to be recognized. In contrast, there was an effect of inversion on dog experts'(dog show judges and breeders with an average of 31 years experience with dogs' appearance) recognition of dog profiles which was comparable to the observed effect of inversion on their recognition of human faces! This result was also found when bird and dog experts were shown bird and dog pictures, and their N170-ERP[20] component was compared: Approximately 164 ms after presentation, objects of expertise (dogs for dog experts; birds for bird experts) can be dissociated from objects from lower expertise categories (Tanaka, 2001b). Thus, based on the schema hypothesis, one would assume that this vast amount of object exposure has resulted in an expert-specific schema that is orientation sensitive because all the exemplars have usually been encountered in the upright position.

Goldstein and Chance did not elaborate on how a schema is used. Nevertheless, the linking element between the results discussed in the previous paragraph might be the processing of configural information in faces: The use of this special class of information could be an essential element of a holistic representation as proposed by Tanaka and Sengco (1997) and might also develop with age as well as the face schema.

8.6 Sensitivity to configuration

Adult face recognition is characterized by a high sensitivity to configural information. For example, Haig (1984) showed for unfamiliar faces that configural alterations produced by changing the distance between facial features are sometimes detected at the visual acuity threshold level. Hosie, Ellis, and Haig (1988) found similar results using familiar faces. Kemp, McManus, and Pigott (1990) used two-tone images and found that the high sensitivity to configural information is reduced in negative or inverted images. While these studies were primarily concerned with the perceptual level, Bruce, Doyle, Dench, and Burton (1991) revealed a specialization for processing configural information at the level of memory processes. When tested, participants had to decide whether faces and houses were identical to the ones presented in a previous block or whether they had been altered configurally. Although the alterations were smaller for faces than for houses, participants were more sensitive in detecting them. Similar to the result of Kemp et al. (1990), this effect diminished when the stimuli were inverted. Leder and Bruce (2000) tested directly whether individual configural elements are represented in memory explicitly. They used a set of 8 faces, each of which differed only in a distinctive local configural feature such as a lowered mouth or a smaller eye-distance. In the test phase, they presented the whole face or the distinctive features in isolation or embedded into an empty head shape. Participants were surprisingly efficient at recognizing faces from the isolated configural elements.

[20] The N170 is a posterior negativity of the event-related potential (ERP) which reflects an early stage of face processing

Moreover, all the experiments in Leder and Bruce (2000) revealed that the processing of configural information was particularly disrupted by inversion. The authors conclude that it is the reliance on configuration that is essential for adult's expertise at processing upright faces.
Thus, based on the review of recent studies, better processing of configural information seems to be applicable for adults rather than children. This is in accordance with findings that the limits of face processing are often accompanied by a disruption of configural rather than other sorts of information. In the next paragraph we describe three effects which are known to be particularly disruptive to adult face processing.

8.7 Testing for limits: The advantage of being upright

The remarkable ability of recognizing faces reliably is highly dependent on orientation. It was already shown in previous chapters how the holistic hypothesis and the use of configural information by adults suggest that orientation is a critical variable. Moreover, to process facial information reliably, a large amount of expertise is required (for a review see Carey, 1992; Chapter 4). Through years of practice, the face recognition system becomes more specialized but at the same time more limited to processing the upright orientation (schema hypothesis). In the following section three effects that illustrate this specialization in upright faces are discussed: the face inversion effect, the Thatcher illusion and the face composite illusion.
In order to investigate whether inversion particularly affects the recognition of faces, Yin (1969) used a forced-choice recognition paradigm with pictures of human faces, airplanes, houses, and stick figures of men in motion as stimuli. In one condition the stimuli were learnt and tested in the upright orientation. Upright faces were recognized better than all the other upright stimuli but were stronger affected by inversion. In another condition the stimuli were learnt in the upright orientation and then tested in the inverted orientation. Generally, when the stimuli had to be recognized in the upside-down position, error rates increased for all stimuli. The interesting finding was that this increase was disproportionately high for faces when compared with the other objects. Whilst faces were recognized best in the upright test condition, performance for inverted faces dropped below the recognition levels of the other object classes. This finding, namely that upside-down faces are disproportionately more difficult to recognize than other inverted objects, has been referred to as the *face inversion effect*. Subsequent replications of Yin's study have refined the initial methodology by comparing faces with stimuli that were equivalent in terms of familiarity, complexity, and psychosexual importance (e.g., Ellis, 1975; Goldstein & Chance, 1981; Scapinello & Yarmey, 1970). Valentine (1988) presented a comprehensive summary of studies investigating the face inversion effect. The review of recent results on holistic and configural processing suggests that the disruption of configural information explains most of the effects of the inversion of faces (Leder & Bruce, 2000).

Another impressive demonstration for the orientation-sensitive nature of face processing comes from a study carried out by Thompson (1980). In a photograph of Margaret Thatcher, he rotated the eyes and mouth within the facial context, which resulted in a grotesque facial expression (see Figure 27 for a demonstration). Interestingly, this strange expression is not perceived when the face is turned upside-down, but is immediately apparent when the face is turned upright. This effect has been referred to as the *Thatcher illusion*. It is clear that this manipulation of the orientation of components alters the form of the eyes and mouth to the point of grotesqueness.

Figure 27. Thatcher illusion. Both inverted pictures look more or less "normal". But when turned upright, the thatcherized version is seen to be highly grotesque. Try it!

Inverting the eyes within the facial context clearly changes the spatial relationship of the parts. Indeed, this alteration has been considered by some authors to produce a change in the configural information (e.g., Bartlett & Searcy, 1993; Diamond & Carey, 1986; Stevenage, 1995).

Young, Hellawell, and Hay (1987) discovered another interesting effect (see Figure 28 for an illustration). They created composite faces by combining the top and bottom half of different faces. If the two halves were aligned and presented upright, a new face resembling each of the two originals seemed to emerge. This made it very difficult to identify the persons from either half. If the top and bottom halves were misaligned horizontally, then the two halves did not fuse spontaneously to create a new face, and the constituent halves remained identifiable. However, when these stimuli were inverted, the constituent halves of the aligned and misaligned displays were equally identifiable. Furthermore, the subjects were significantly faster at naming the constituent halves in inverted composites than in upright composites. Young et al. (1987) have argued that it is the new configuration in the composite face, which makes the identification of the parts difficult. Thus again we have evidence that an effect specific for upright faces might be due to the use of configural information in upright faces and the disruption of this in upside-down faces.

Figure 28. Aligned and misaligned halves of different identities (here two of the authors). When upright (as above), a new identity seems to emerge from the aligned composites (left), which makes it more difficult to extract the original identities. This does not occur for the misaligned composite face (right). When viewed upside-down, the original identities can be extracted easily from both pictures.

Concerning the developmental course, Cashon and Cohen (2001) showed that 7-month-old infants process composites from outer and inner features as one face. This may be taken as evidence for a kind of configural processing, which is in accordance with Tanaka, Kay, Grinnell, Stansfield, and Szechter (1998) who found that 6-year- olds showed the same whole-to-part superiority effects as adults. Carey and Diamond (1994) also found that adult-like composite effects emerge at the age of 6 while configural processing (indicated by inversion effects) develops continually until adulthood. Recently, Mondloch, Le Grand, and Maurer (2002) showed that configural processing develops later than featural or component processing and that it may still develop after the age of ten.

8.8 Component configural hypothesis

While numerous studies have been presented which stress the importance of configural processing, it is not yet clear how different features are combined to form a representation of faces in memory. According to the component configural hypothesis, component and configural information is processed separately, and configural processing is much more affected by changes of orientation than the processing of components. There is a large number of studies in favor of this view. The first demonstration of a differential effect of inversion on the processing of component and configural information was provided by Sergent (1984). She used pairs of faces where either the eyes or facial contour (change of component information) or the internal spacing of components (change of configural information) were mismatched. A multidimensional scaling technique for the analysis of dissimilarity judgments, and regression analyses on reaction times revealed that configural and component information were used for upright faces. In contrast, there was no evidence that subjects made use of configural information when faces were inverted. It should be noted, however, that Sergent (1984) used schematic faces which could make it difficult to generalize this result to the processing of real faces. Nevertheless, similar results were found by Searcy and Bartlett (1996), who used color photographs of faces in which configural changes had been induced by moving the eyes and mouth up or down, and manipulation of the component information had been achieved by changing the color of the pupils and teeth or by shortening and elongating the teeth. In line with Sergent’s (1984) results, a grotesqueness-rating task and a simultaneous paired-comparison task provided further evidence for the view that inversion is particularly disruptive to the processing of configural information. Leder and Bruce (1998) manipulated the distinctiveness of either components or configural features directly and showed how both make upright faces easier to recognize. When faces were presented upside-down, the effects of distinctiveness based on configural features vanished in nearly all conditions.

Another demonstration of the differential effects of orientation on the processing of component and configural information was provided in chapter 2. Using a sequential same-different matching task it was shown that the detection of component changes (eyes and mouth replaced) was relatively invariant to planar rotations. In contrast, rotation had a detrimental effect upon the detection of configural changes that were induced by increasing the distance between the eyes and the eyes and mouth (Figure 29). Interestingly, the effect of rotation on configural processing was nonlinear; most errors were found at intermediate angles of rotation between upright and inverted orientations, i.e., at 90° – 120°. Similarly, Murray, Yong, and Rhodes (2000) found a discontinuity in the function relating bizarreness to a rotation of between 90° and 120° which was found for Thatcher faces and faces in which configural changes were induced by changing the relative position of the eyes and mouth. The bizarreness ratings of unaltered or component-distorted faces (teeth blackened and eyes whitened)

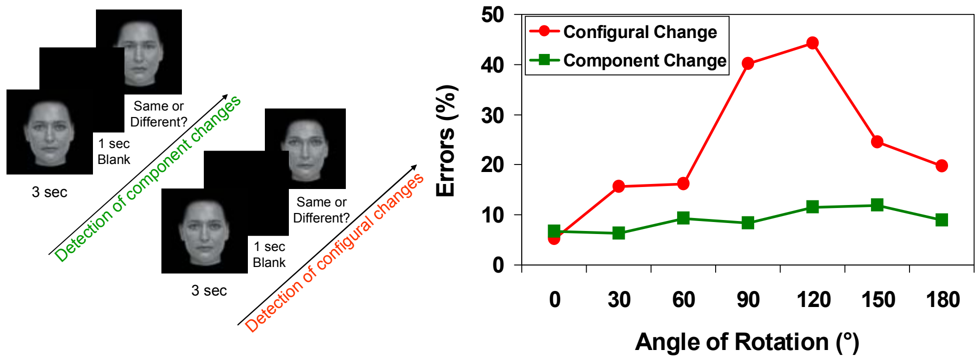

Figure 29. Study from chapter 2 (Experiment 1). Left: The detection of component and configural changes was tested using a sequential same-different matching task in separate experiments. Right: Whereas the identification of component changes was almost unaffected by rotation, the detection of configural changes was strongly impaired.

showed only a linear trend. Leder and Bruce (2000, Experiment 5) compared directly whether configurations are also accessible when, at the same time, components vary from face to face: the isolated configurations, though composed of components which they shared with other faces, were recognized and showed inversion effects. To show directly that configural information is processed differentially in upright as compared to inverted faces, Leder, Candrian, Huber, and Bruce (2001) used a sequential comparison task. Participants saw two faces sequentially which differed in interocular eye-distance only. The task was to decide for each pair of faces which face had the larger interocular eye-distance. The judgments were more accurate when the faces were presented upright, and the decrement in accuracy in the inverted condition was independent of the size of the surrounding context (e.g., whether the nose or the mouth and nose were added).

One possible caveat of the studies that investigated the processing of component and configural information by replacing or altering facial features is that this type of manipulation often changes the holistic aspects of the face and is difficult to carry out selectively. For example, replacing the nose (component change) can change the distance between the contours of the nose and the mouth and thus alter the configural information (Leder & Bruce, 1998; 2000). The same applies to configural changes when they are carried out by altering the relative position of the components. For example, moving the eyes apart (configural change) can lead to an increase in size of the bridge of the nose, i.e., a component change (see Leder et al., 2001).

Problems like these can be avoided by using scrambling and blurring procedures to eliminate configural and component information separately. The techniques used in chapter 4 extend previous research by ensuring that scrambling and blurring effectively eliminate configural and component information separately. Furthermore, in contrast to

previous studies, the same faces in separate experiments on unfamiliar and familiar face recognition were used to avoid potential confounds with familiarity (Figure 4).
In Experiment 7, unfamiliar face recognition was studied. In the first condition it was shown that previously learnt intact faces could be recognized even when they were scrambled into constituent parts. This result challenges the assumption of purely holistic processing according to Farah et al. (1995) and suggests that facial features or components are encoded and stored explicitly. In a second condition, the blur level was determined that made the scrambled versions impossible to recognize. This blur level was then applied to whole faces in order to create configural versions that by definition did not contain local featural information. These configural versions of previously learnt intact faces could be recognized reliably. This result suggests that separate representations exist for component and configural information. Familiar face recognition was investigated in Experiment 8 by running the same conditions with participants who knew the target faces (all distractor faces were unfamiliar to the participants). Component and configural recognition was better when the faces were familiar, but there was no qualitative shift in processing strategy since there was no interaction between familiarity and condition (Figure 4).

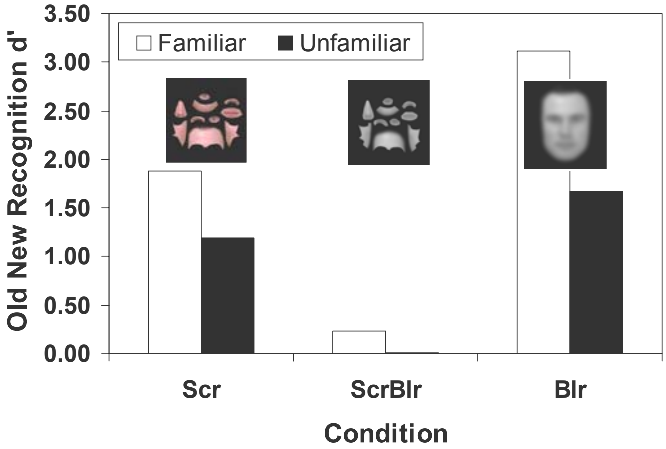

Figure 30. Recognition performance in unfamiliar and familiar face recognition across three different conditions at test. Scr: scrambled, ScrBlr: scrambled and blurred, Blr: blurred. (Adapted from Schwaninger, Lobmaier, & Collishaw, 2002)

In sum, there is converging evidence in favor of the view that separate representations for component and configural information exist which are relevant for the recognition of familiar and unfamiliar faces. Whereas component information is not very orientation-sensitive, configural information is difficult to recover when faces are rotated.

8.9 An integrative view of face recognition

Everyday object recognition is often a matter of discriminating between quite heterogeneous object *classes* that differ with regard to their global shape, parts and other distinctive features such as color or texture. In contrast, face recognition relies on the discrimination of *exemplars* of a very homogenous category. All faces share the same basic parts in the same basic arrangement. In each face the eyes are above the nose which is located above the mouth. Therefore, reliable face recognition relies on the detection of subtle featural and configural differences, which needs years of experience. Since faces are usually seen upright, this learning must become more and more restricted to the upright orientation. A strong dependency on orientation is the consequence for objects that are usually perceived in one specific orientation. Since effects of rotation and inversion are much more detrimental for faces than for basic

level object recognition, a certain type of information must be more relevant for faces. According to certain authors, expert face recognition is characterized by holistic processing (e.g., Biederman & Kalocsai, 1997; Farah et al., 1995; Tanaka & Farah, 1993). Farah et al. (1995) answer the question "Why is face recognition so orientation sensitive?" in the following way: "Face perception is holistic and the perception of holistically represented complex patterns is orientation sensitive." (p. 633). According to Rock (1973, 1974, 1988), rotated faces overtax an orientation normalization mechanism, which makes it impossible to match them against stored upright memory representations. Rotated faces can only be processed by their components, and configural information is hard to recover. This would explain why effects of rotation are much smaller for component as opposed to configural changes (Leder & Bruce, 1998, 2000; Schwaninger & Mast, 1999). At the same time, these results challenge a purely holistic view of face processing which assumes that explicit representations of facial parts do not exist. The results obtained in Experiment 7 and 8 (see chapter 4) offer further evidence against such a purely holistic view. They revealed that facial components and configural information are encoded and stored explicitly, both in unfamiliar and familiar face recognition, when faces are upright.
In order to integrate the different hypotheses discussed in this thesis the model depicted in Figure 31 is useful. All pictorial aspects of a face are contained in the pictorial metric input representation which is presumably correlated with activation in primary

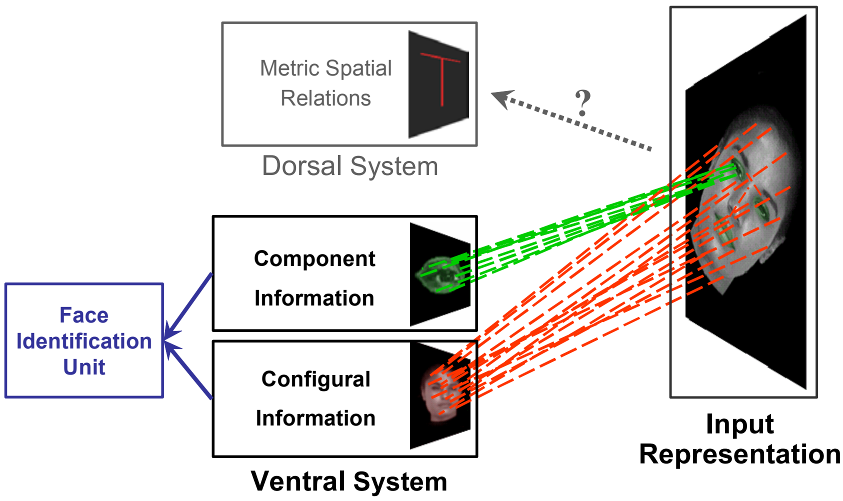

Figure 31. Integrative model of face processing. Facial information is encoded in a metric input representation that contains all the features we perceive in faces. Information of local features and relations between them is extracted in order to activate component and configural representations in the ventral stream. The outputs of these representations converge towards the same face identification units. Whether dorsal processing is relevant for processing metric spatial relations in faces such as the eye-mouth or the inter-eye distance remains to be investigated.

visual areas. Based on years of expertise, neural networks are trained to extract specific information in order to activate component and configural representations in the ventral visual stream. The output of these representations converges towards the same identification units. These units are holistic in the sense that they integrate component and configural information. Note that this concept of holistic differs from the original definition of Tanaka and Farah (1993) and Farah et al. (1995). In their view, holistic means that parts are not represented explicitly. In contrast, according the integrative model proposed here, holistic processing implies that component and configural information are first encoded separately and then integrated into a holistic representation. This concept of holistic is fully compatible with the results from Schwaninger et al. (2002) and Leder et al. (2001) who showed that featural and configural information is encoded explicitly. Moreover, the integrative definition of holistic proposed here is consistent with the results of Tanaka and Sengco (1997) and Rhodes et al. (1993) which imply that in normal (upright) face processing, component and configural information is combined into a single holistic face representation.

Adult face recognition is characterized by the processing of configural information and by the fact that faces are quite hard to recognize when they are rotated substantially from the upright position. In the model this can be explained in the following way: When faces are rotated, the pictorial information in the input representation is changed remarkably. As a consequence, the component and configural representations which have been learnt based on exposure to upright faces, cannot be activated well enough to allow reliable recognition.

Rotated faces overtax orientation normalization mechanisms so that they have to be processed by their components (Rock, 1973, 1974, 1988). As pointed out by Valentine and Bruce (1988), this implies that information about the spatial relationship of components (configural information) is hard to recover. Consequently, the processing of configural information is much more affected by rotation or inversion than the processing of component information. Since face recognition relies heavily on processing configurations, the inversion effect is in disproportion to that of other objects (Yin, 1969). This is the deeper answer to the question "Why is face recognition so orientation sensitive?"

The integrative model proposed here also offers an explanation for the Thatcher illusion and the composite face illusion. Thatcherizing a face, i.e., inverting the eyes and mouth within an upright face, results in a strange activation pattern of component and configural representations. Consequently, the face looks very bizarre. When a thatcherized face is inverted, the activation of configural representations is strongly impaired due to the limitation in capacity of an orientation normalization mechanism. Consequently, the strange activation pattern of configural representations is reduced and the bizarre perception vanishes. Moreover, in an inverted Thatcher face the components themselves are in the correct orientation which results in a relatively normal activation of component representations. Consequently, inverted Thatcher faces appear relatively normal (Rock, 1988). Finally, the composite face illusion can be explained by similar reasoning. Aligned upright face composites contain new configural information resulting in a new perceived identity. Inverting the aligned composites reduces the availability of configural information and it is easier to access the two different face identification units based on the component information alone.

In short, the model proposed here allows the integration of the component configural hypothesis and holistic aspects of face processing. It explains striking perceptual effects such as the Thatcher illusion and the composite face illusion. Most importantly, it provides an integrative basis for understanding special characteristics of adult face recognition such as the specialization in upright faces and the sensitivity to configural information.

8.10 References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54-75.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*(3)*,* 281-316.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*(2)*,* 115-147.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London, B, 352*, 1203-1219.

Brown, R. (1958). How shall a thing be called? Psychological Review, 65, 14-21.

Bruce, V. (1988). *Recognising faces*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*, 109-144.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London*, *335*, 95-103.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*, 312-314.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1*, 253-274.

Cashon, C., & Cohen, L. (2001). Do 7-month-old infants independent features of facial configurations? *Infant and child development, 10*, 83-92.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29*, 893-910.

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes, *Acta Psychologica, 73* 1990 225–243.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Ellis, A. W. (1992). Cognitive mechanisms of face processing. In V. Bruce, A. Cowey, A. W. Ellis, & D. P. Perrett (Eds.), *Processing the facial image* (pp. 113-119). New York, NY: Clarendon Press.

Ellis, H. (1975). Recognizing faces. *British Journal of Psychology*, *66*, 409-426.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 628-634.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105*, 47-59.

Goldstein, A. G., & Chance, J. E. (1981). Laboratory studies of face recognition. In G. M. Davies, H. D. Ellis, & J. W. Sheperd (Eds.), *Perceiving and remembering faces*. London, England: Academic Press.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception, 13(5),* 505-12.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62,* 201-222.

Hosie, J. A., Ellis, H. D., & Haig, N. D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception*, *17*, 461-474.

Johnston, R.A., & Ellis, H.D. (1995). The development of face recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition, explorations in face space* (pp. 1-23). London, New York: Routledge.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19*, 531-543.
Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology, 51A*, 449-473.
Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, 53A*, 513-536.
Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30*, 73-83.
Macho S., & Leder H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *Journal of Experimental Psychology: Human Perception and Performance, 24(5),* 1486-1500.
Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural processing develops more slowly than feature face processing. *Perception, 31*, 553-566.
Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11*, 498-502.
Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition, 47*, 25-57.
Rock, I. (1973). *Orientation and form*. New York: Academic Press.
Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230*, 78-85.
Rock, I. (1988). On Thompson's inverted-face phenomenon (Research Note). *Perception, 17*, 815-817.
Scapinello, F. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21*, 329-331.
Schwaninger, A., & Mast, F. (1999). Why is face recognition so orientation-sensitive? Psychophysical evidence for an integrative model. *Perception (Suppl.), 28*, 116.
Schwaninger, A., Lobmaier, J., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science, 2525,* 643-650.
Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information of faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904-915.
Sergent, J. (1984). An investigation into component and configurational processes underlying face recognition. *British Journal of Psychology , 75,* 221-242.
Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *Journal of Experimental Psychology: Human Perception and Performance, 11(6)*, 846-61.
Stevenage, S. V. (1995). Expertise and the caricature advantage. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition, explorations in face space* (pp. 24-46). London, England: Routledge.
Tanaka, J. W. (2001a). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130*, 534-543.
Tanaka, J. W. (2001b). A neural basis for expert object recognition. *Psychological Science, 12*, 43-47.
Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*, 225-245.
Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*, 583-592.
Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*, 479-496.

Thompson, P. (1980). Margaret Thatcher -- A new illusion. *Perception*, *9*, 483-484.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36(12),* 1761-1771.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory and Cognition, 16,* 556-566.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Young, A. W., Hellawell, D. J., & Hay, D. C. (1987). Configural information in face perception. *Perception, 16*, 747-759.

9 Danksagung

Die Wahrnehmung und Wiedererkennung von Gesichtern ist ein ebenso relevantes wie spannendes Thema. Bei den verschiedenen Experimenten dieser Arbeit wurde ich durch hervorragende Studenten unterstützt. Bei Franziska Hofer und Stefan Ryf bedanke ich mich für die tolle Zusammenarbeit bei den Experimenten in Kapitel 3. Die Studie in Kapitel 4 wurde zusammen mit Janek Lobmaier durchgeführt. Sandra Schumacher hat wertvolle Beiträge geleistet bei den Experimenten in Kapitel 6. Ich danke ganz herzlich diesen ausgezeichneten Nachwuchskräften, welche bei ihrer Lizentiatsarbeit alle eine sehr gute Eignung für das wissenschaftliche Arbeiten ausgewiesen haben.

Wie bei anderen modernen Forschungsgebieten ist eine Zusammenarbeit mit anderen Forschern essentiell um die komplexen Prozesse unseres kognitiven Apparates besser zu verstehen. Bei verschiedenen Kollegen möchte ich mich für wertvolle Diskussionen und Kollaborationen bedanken. Zusammen mit Stefan Collishaw (University of Sussex) habe ich die Studien in Kapitel 5 durchgeführt. Die Arbeiten in Kapitel 6 und 7 sind in Zusammenarbeit mit Christian Wallraven (MPI Tübingen) entstanden. Bei Helmut Leder und Claus Christian Carbon (Universität Berlin) bedanke ich mich für ihre Beiträge und Feedback zu Kapitel 8.

Meinen Betreuern möchte ich auch besonderen Dank aussprechen. Für die Arbeiten in Kapitel 2 bin ich Fred Mast dankbar, welcher mich auch nach der Lizentiatsarbeit am Anfang meiner Doktorarbeit betreut hatte. Wolfgang Marx bin ich sehr dankbar für die wertvolle Unterstützung in den Jahren als Student und Doktorand. Hans Heinrich Bülthoff (MPI Tübingen) danke ich für wichtiges Feedback im Bezug auf experimentelles Design, sowie dem Schreiben und Publizieren von Artikeln. Den Arbeitskollegen an der Universität Zürich und am MPI Tübingen danke ich für wertvolle Diskussionen und eine gute Zusammenarbeit auch bei administrativen Dingen.

Schon in den ersten Jahren meiner Doktorarbeit habe ich auch angewandte Projekte am Flughafen Zürich begonnen. Für die tolle Zusammenarbeit bei diesen Projekten bedanke ich mich ganz herzlich bei den betreffenden Mitarbeitern und Kaderpersonen der Kantonspolizei Zürich, Flughafenpolizei und bei der Flughafen Zürich AG für ihre Unterstützung.

Besonders danken möchte ich meinen Eltern, meinem Bruder, meiner Freundin und meinen Freunden für die emotionale Unterstützung vor und während der Doktorarbeit.

Last but not least danke ich allen Versuchspersonen, welche am einen oder anderen Experiment mitgemacht haben.

10 Lebenslauf Adrian Schwaninger

Persönliche Informationen

Staatsangehörigkeit: Schweiz
Familienstand: ledig
Alter: 32
Geburtsort: Zürich

Ausbildung

1999-2003	Promotionsstudium (Universität Zürich) Dissertation Mai 2003. Titel: Perception and representation of faces
1997	Praktikum am Institut für Psychobiologie der Universität Autonoma, Madrid (2 Monate)
1991-1999	Studium der Psychologie, Neuropsychologie, Informatik und Neurophysiologie an der Universität Zürich Lizentiat Juni 1999. Titel: Processing of component and configural information and mental rotation of faces
1990	Sprachaufenthalt am LCP Institute Covina, Kalifornien (3 Monate)
1984-1990	Kantonsschule Winterthur (Matura Typus B)

Berufserfahrung

2000-2002	Koordinator des Graduate Program Cognitive Psychology Switzerland [http://vision.unibe.ch/graduate/].
Seit 2001	Doktorand am Max-Planck-Institut für Biologische Kybernetik, Tübingen, Deutschland
Seit 2000	Projektleitung am Flughafen Zürich in Zusammenarbeit mit der Kantonspolizei Zürich, Flughafenpolizei: Steigerung der Sicherheit und Effizienz bei der Erkennung verbotener Gegenstände in Röntgenbildern bei Sicherheitskontrollen
Seit 1999	Lehrbeauftragter an der Universität Zürich und ETH Zürich
Seit 1998	Assistent an der Abteilung Allgemeine Psychologie am Psychologischen Institut der Universität Zürich, Leitung der Projektgruppe Visuelle Kognition
1996-2000	Teilzeitanstellung als Kursleiter bei IBM Schweiz
1995-1998	Semesterassistent am Psychologischen Institut der Universität Zürich
1992-1998	Teilzeitanstellung bei Securitas AG Zürich